DOCUMENT RESUME

ED 304 459 TM 012 831

AUTHOR Lawrence, Ida M.; And Others
TITLE Differential Item Functioning for Males and Females on SAT-Verbal Reading Subscore Items. College Board Report No. 88-4.
INSTITUTION College Entrance Examination Board, New York, N.Y.; Educational Testing Service, Princeton, N.J.
REPORT NO ETS-RR-88-10
PUB DATE 88
NOTE 62p.
AVAILABLE FROM College Board Publications, P.O. Box 886, New York, NY 10101-0886 ($6.00).
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS *College Entrance Examinations; Higher Education; High Schools; High School Students; *Item Analysis; Reading Comprehension; *Reading Tests; Science Tests; Sentences; *Sex Differences; *Verbal Tests; Writing Evaluation
IDENTIFIERS *Differential Item Performance; *Scholastic Aptitude Test; Sentence Completion Test

ABSTRACT

The reading comprehension (RC) and sentence completion (SC) items from four forms of the Scholastic Aptitude Test (SAT) verbal sections were examined for differential item functioning (DIF) between male and female examinees. Forms 1, 2, 3, and 4 of the SAT-verbal were administered in January 1983, January 1984, October 1984, and June 1986, respectively. The sample included a total of 83,493 male and 92,482 female secondary school juniors and seniors. In general, minimal DIF was observed. However, among the relatively few items exhibiting substantial DIF, two trends related to science content were noted. An important factor that appears to be connected to DIF on RC items is the distinction between technical and non-technical reading passage material. There seems to be an interaction between the technical science factor and the passage location, although confirmation of this finding requires further study. The main factor that appears related to DIF on SC items is the distinction between non-science or surface science references and true science references. Items with true science references were more difficult for females. Findings also indicate the confounded nature of multiple factors thought to contribute to DIF. There is a need for experimental research in which item features and passage features can be controlled and counterbalanced in variable sections of the SAT. Thirteen data tables and 11 graphs are presented. Selected items and analyses and 136 graphs are appended. (TJH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED304459

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

PATRICIA K. HENDEL

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

College Board Report

No. 88-4

Differential Item Functioning for Males and Females on SAT® -Verbal Reading Subscore Items

Ida M. Lawrence
W. Edward Curley
Frederick J. McHale

Differential Item Functioning for Males and Females on SAT®-Verbal Reading Subscore Items

Ida M. Lawrence
W. Edward Curley
Frederick J. McHale

College Board Report No. 88-4

ETS RR No. 88-10

College Entrance Examination Board, New York, 1988

Ida M. Lawrence is a measurement statistician at Educational Testing Service, Princeton, New Jersey

W. Edward Curley is a test development examiner at Educational Testing Service, Princeton, New Jersey

Frederick J. McHale is a senior measurement statistician at Educational Testing Service, Princeton, New Jersey

Acknowledgments

The authors wish to thank Behroz Maneckshana for assisting with the data processing; Edwin Blew, Susan Bryce, and Karen Carroll for producing the graphics; and William Angoff, Neil Dorans, and Alicia Schmitt for reviewing the earlier drafts of this report.

Researchers are encouraged to express freely their professional judgment. Therefore, points of view or opinions stated in College Board Reports do not necessarily represent official College Board position or policy.

The College Board is a nonprofit membership organization that provides tests and other educational services for students, schools, and colleges. The membership is composed of more than 2,500 colleges, schools, school systems, and education associations. Representatives of the members serve on the Board of Trustees and advisory councils and committees that consider the programs of the College Board and participate in the determination of its policies and activities.

Additional copies of this report may be obtained from College Board Publications, Box 886, New York, New York 10101-0886. The price is $6.

Copyright © 1988 by College Entrance Examination Board. All rights reserved.
The SAT questions in Appendix A are reprinted by permission of Educational Testing Service, the copyright owner.
College Board, Scholastic Aptitude Test, SAT, and the acorn logo are registered trademarks of the College Entrance Examination Board.
Printed in the United States of America.

CONTENTS

Abstract 1

The Study 1

Method 2
Data Source 2
Procedures 2
Item Classification 3

Results and Discussion 4
Distribution of DIF for the Four Verbal Item Types 4
Reading Comprehension Items 4
Distractor Analyses for Reading Comprehension Items 9
Sentence Completion Items 14
Distractor Analyses for Sentence Completion Items 16
Summary of Findings for Exploratory Analyses of Forms 1, 2, and 3 17
Confirmatory Analyses for Form 4 17

Conclusions 22

References 23

Appendix A: Selected Reading Comprehension Items, Distractor Analyses, and Empirical Characteristic Functions 24

Appendix B: Selected Sentence Completion Items, Distractor Analyses, and Empirical Characteristic Functions 47

Figures

1. Standardized Percentage DIF Values for Males and Females on SAT Reading Comprehension Items: Form 1 8

2. Standardized Percentage DIF Values for Males and Females on SAT Reading Comprehension Items: Form 2 9

3. Standardized Percentage DIF Values for Males and Females on SAT Reading Comprehension Items: Form 3 9

4. Standardized Percentage DIF Values for Males and Females, by Item Difficulty: Form 1 11

5. Standardized Percentage DIF Values for Males and Females, by Item Difficulty: Form 2 12

6. Standardized Percentage DIF Values for Males and Females, by Item Difficulty: Form 3 12

7. Standardized Percentage DIF Values for Males and Females on SAT Sentence Completion Items: Form 1 15

8. Standardized Percentage DIF Values for Males and Females on SAT Sentence Completion Items: Form 2 15

9. Standardized Percentage DIF Values for Males and Females on SAT Sentence Completion Items: Form 3 16

10. Standardized Percentage DIF Values for Males and Females on SAT Reading Comprehension Items: Form 4 20

11. Standardized Percentage DIF Values for Males and Females on SAT Sentence Completion Items: Form 4 20

A.1. Empirical Characteristic Functions for Reading Comprehension Item 22 in Form 1, Verbal 1 30

A.2. Empirical Characteristic Functions for Reading Comprehension Item 23 in Form 1, Verbal 1 32

A.3. Empirical Characteristic Functions for Reading Comprehension Item 25 in Form 1, Verbal 1 34

A.4. Empirical Characteristic Functions for Reading Comprehension Item 25 in Form 3, Verbal 1 36

A.5. Empirical Characteristic Functions for Reading Comprehension Item 27 in Form 3, Verbal 2 38

A.6. Empirical Characteristic Functions for Reading Comprehension Item 28 in Form 3, Verbal 2 40

A.7. Empirical Characteristic Functions for Reading Comprehension Item 29 in Form 4, Verbal 1 42

A.8. Empirical Characteristic Functions for Reading Comprehension Item 27 in Form 4, Verbal 2 44

B.1. Empirical Characteristic Functions for Sentence Completion Item 14 in Form 1, Verbal 2 48

B.2. Empirical Characteristic Functions for Sentence Completion Item 18 in Form 3, Verbal 1 50

B.3. Empirical Characteristic Functions for Sentence Completion Item 20 in Form 4, Verbal 1 52

B.4. Empirical Characteristic Functions for Sentence Completion Item 11 in Form 4, Verbal 2 54

Tables

1. Summary of Sample Sizes and Descriptive Statistics 2

2. Distribution of Standardized P-Differences between Male and Female Examinees, by Item Type: Form 1 5

3. Distribution of Standardized P-Differences between Male and Female Examinees, by Item Type: Form 2 6

4. Distribution of Standardized P-Differences between Male and Female Examinees, by Item Type: Form 3 7

5. Distribution of Standardized P-Differences between Male and Female Examinees, by Length of Reading Passage 13

6. Pictorial Diagram of Standardized P-Differences between Male and Female Examinees on Reading Comprehension Items, by Science Content 13

7. Difficulty Distributions in Equated Delta Units, by Form and Science Content: Reading Comprehension Items 14

8. Difficulty Distributions in Equated Delta Units, by Form and Science Content: Sentence Completion Items 17

9. Distribution of Standardized P-Differences between Male and Female Examinees on Sentence Completion Items: Factors Identified in Previous Research 18

10. Pictorial Diagram of Standardized P-Differences between Male and Female Examinees on Sentence Completion Items, by Science Content 18

11. Distribution of Standardized P-Differences between Male and Female Examinees, by Item Type: Form 4 19

12. Pictorial Diagram of Standardized P-Differences between Male and Female Examinees on Reading Comprehension Items, by Science Content: Form 4 21

13. Pictorial Diagram of Standardized P-Differences between Male and Female Examinees on Sentence Completion Items, by Science Content: Form 4 21

A.1. Reading Comprehension Text for SAT Form 1, Verbal 1, Items 21–25, with Standardized Proportion Difference between Males and Females 24

A.2. Reading Comprehension Text for SAT Form 3, Verbal 1, Items 21–25, with Standardized Proportion Difference between Males and Females 25

A.3. Reading Comprehension Text for SAT Form 3, Verbal 2, Items 26–28, with Standardized Proportion Difference between Males and Females 26

A.4. Reading Comprehension Text for SAT Form 4, Verbal 1, Items 26–30, with Standardized Proportion Difference between Males and Females 27

A.5. Reading Comprehension Text for SAT Form 4, Verbal 2, Items 26–28, with Standardized Proportion Difference between Males and Females 29

B.1. Examples of Sentence Completion Items with High Values of STPDIF 47

ABSTRACT

The Reading Comprehension and Sentence Completion items from four forms of the SAT-verbal sections were examined for differential item functioning (DIF) between male and female test-takers. In general, minimal DIF was observed for the items analyzed in this investigation. However, among the relatively few items exhibiting substantial DIF, two trends related to science content were noted.

An important factor that appears to be connected to DIF on Reading Comprehension items is the distinction between technical and nontechnical reading passage material. Items associated with passages containing technical (as opposed to historical or philosophical) science material are generally more difficult for female examinees. This trend emerged for three of the four forms studied. Data from the fourth form suggested that there may be an interaction between the technical science factor and the passage location. Confirmation of this finding will require a follow-up study.

The main factor that appears to be related to DIF on Sentence Completion items is the distinction between nonscience or surface science references and true science references. Items containing true science references tend to be more difficult for females.

Because the majority of items in the Reading subscore exhibited minimal DIF, the conclusions described in this report are based on limited data. Findings from this study also indicate the confounded nature of multiple factors thought to contribute to DIF. There is a need for experimental research in which item features and passage features can be controlled and counterbalanced in variable sections of the SAT.

THE STUDY

Much attention has been focused on differences in performance by men and women taking the Scholastic Aptitude Test (SAT). In an attempt to understand the nature of some of these sex differences, we looked at a subset of SAT-verbal items—namely, the items constituting the Reading subscore. The goal of this study was to determine whether differential item functioning (DIF) between males and females can systematically be attributed to certain factors in the item content.

Statistical data compiled by Educational Testing Service (ETS) in the *Summary of ATP Scores 1965–66 through 1984–85* indicate that since 1972, there has been a larger average decline in SAT-verbal scores for women than there has been for men. It appears that the Reading subscore has contributed more to the change in the male/female difference than has the Vocabulary subscore. Since the introduction of SAT-verbal subscore reporting in 1975, there has been a relatively constant difference between males and females on the Vocabulary subscore. However, for the Reading subscore, mean scores for males and females began to diverge in 1978, and in 1981 the gap widened to yield a difference similar to that on the Vocabulary subscore. This gap in Reading subscores has persisted.

This study was initiated because of concerns that slight changes in the content and structure of the SAT-verbal items may have contributed to a larger score decline for women than for men. Some minor changes in the test involved both the content and the number of reading passages and items. Before 1974, the SAT-verbal sections consisted of 90 items administered in 75 minutes. Included were seven reading passages, usually 375 to 475 words each; five items followed each of the passages for a total of 35 Reading Comprehension questions. Content specifications called for each form of the test to include two science passages. In 1974, with the introduction of the Test of Standard Written English (TSWE), testing time for the SAT-verbal sections was reduced to 60 minutes, and the total number of verbal items was decreased to 85. Five 375-to-475-word passages, each followed by five items, yielded a total of 25 Reading Comprehension questions. Only one science passage was called for in the modified content specifications. Finally, in 1978, shorter reading passages (usually 175 to 275 words each) were introduced into the verbal sections of the SAT. The total number of Reading Comprehension items remained at 25, but six passages rather than five were included in each form (three longer passages plus three shorter ones), and the shorter passages were followed by as few as two to three questions. New content specifications for the item type, which have remained unchanged to the present, called for a return to two science passages as in the pre-1974 test.

Given this change in the number and the content of Reading Comprehension passages and items, and the increased difference in Reading scores between males and females that began in 1978, a natural focus of study was the Reading Comprehension item type. Differential item functioning between males and females on the Sentence Completion item type also warranted attention because the Reading subscore comprises both types.

Reading Comprehension items pose a unique problem for studying DIF, because sets of individual items are related to a single reading passage. If the particular reading passage contains content disadvantageous to a particular subgroup, differential functioning may be found for all of the items pertaining to that passage, or for only some of the items. Therefore, in addition to examining individual items for the presence of DIF, we considered the functioning of items within a passage as well, by looking at DIF over the subset of items that belong to a given reading passage.

Previous research on DIF for male and female SAT

candidates has focused on item type (Dorans and Kulick 1983) and on traditional test development content specification categories (Wendler and Carlton 1987) to generate hypotheses to explain why certain SAT-verbal items function differentially. An attempt was made in this report to go beyond the traditional item classifications by focusing on additional content features of the items or reading passages that might explain DIF. This research is viewed as a preliminary effort to learn more about the scope of DIF for females on Reading items and to gain a better understanding of the differential performance of males and females on particular items.

The investigation was conducted in two stages. In the first phase, Reading Comprehension and Sentence Completion items from three SAT forms were examined for DIF. Several possible sources of DIF were identified. These findings were used to generate a set of hypotheses that were then tested on a fourth form.

METHOD

Data Source

Four forms of the SAT-verbal sections were examined for differential item functioning between male and female test-takers. Referred to in this report as Forms 1, 2, 3, and 4, these forms were administered in January 1983, January 1984, October 1984, and June 1986, respectively. The data for the study were the responses of male and female examinees on the Reading Comprehension and Sentence Completion items that appeared in the operational sections of these forms. For each test form, samples of male and female examinees were restricted to secondary school juniors and seniors who had indicated on the Student Descriptive Questionnaire (SDQ) that English was their best language. A random sample from the restricted population was available and analyzed for Form 1, while the entire restricted populations were analyzed for Forms 2, 3, and 4.

In Table 1 sample sizes for Forms 1, 2, 3, and 4 are displayed along with information concerning relative subgroup and population ability. Note that raw score summary statistics for the different forms are not on the same scale; form-to-form comparisons, therefore, cannot be made. Within each test form, however, raw score means may be used to compare the relative ability of the two subgroups. For Forms 1, 2, 3 and 4, the raw score differences, male minus female mean scores, on SAT-verbal items are .14, .12, .14, and .17 in standard deviation (SD) units, respectively, indicating that as a group, males perform better than females do on the verbal portions of the SAT. At the bottom of Table 1, the scaled score summary statistics indicate that the test forms examined in this study were administered to two relatively low-ability groups and two relatively high-ability groups,

Table 1. Summary of Sample Sizes and Descriptive Statistics

Sampled Examinees for DIF Analyses[a]	*Form* 1	2	3	4
Males				
Sample *N*	6,329	25,986	14,529	36,649
Raw Score Mean	33.18	34.48	40.49	39.18
Raw Score SD	16.61	16.34	16.24	16.03
Females				
Sample *N*	6,712	27,158	17,822	40,790
Raw Score Mean	30.85	32.58	38.30	36.55
Raw Score SD	15.98	15.93	16.01	15.58
Information on Representative Samples of Juniors and Seniors[b]				
Scaled Score Mean	407	403	433	434
Scaled Score SD	104	109	108	104
Test Reliability	.920	.925	.928	.923

a. Samples were restricted to secondary school juniors and seniors for whom English is the best language.

b. Samples of juniors and seniors as selected for SAT Test Analysis Reports.

where performance on SAT-verbal items is reported on the College Board scale, which ranges from 200 to 800.

Procedures

The standardization method developed by Dorans and Kulick (1983, 1986) was used in this study to assess DIF between male and female examinees on items constituting the Reading subscore. The standardization method identifies differences in item functioning between a focal group and a reference group, after controlling for differences in ability level. For the four forms analyzed in this investigation, items were binary-scored (right/wrong), and the matching criterion was the total formula score on the SAT-verbal items. The focal group was defined as the female group.

In the standardization procedure a weighting function is used that reflects the relative frequency of focal group examinees at each score level. Differences in item performance between the reference group and the focal group at each score level are weighted and then summed across score levels to arrive at a summary index of DIF. In this report the summary index is expressed either as a standardized difference in percentage correct (Figures 1 to 11) or as a standardized difference in proportion correct (Tables 2 to 6 and 9 to 13). The acronym STPDIF is used for both. For each of the four test forms, DIF statistics for the 15 Sentence Completion items and 25 Reading Comprehension items were computed.

Distributions of STPDIF values for individual items, rather than summary statistics, were used to identify factors that may contribute to DIF. This approach

was chosen because, within a given test form, most of the content classification categories include only a small number of items (generally fewer than 10). Summary statistics are bound to be unstable when the numbers of items per classification are as few as those encountered in this study, and in the event that there are one or two extreme values, such statistics may even provide misleading information.

The distributions of STPDIF values for each test form, broken down by the two item types, were examined for the most extreme differentially functioning items. Within these distributions, clear outliers were scrutinized with respect to a variety of factors to determine whether it was possible to explain why such items exhibit DIF. Careful attention was directed at those items that appear most discrepant relative to the distribution of other items in the test form. For most items, it should be noted, STPDIF values are distributed between −.05 and .05.

In addition to examining DIF with respect to the keyed item response, we conducted distractor analyses, again using scores on SAT-verbal items as the matching criterion. As part of the distractor analyses, DIF statistics were calculated in two ways, using (1) all examinees who took the test and (2) all examinees who reached the item. Distractor analyses make it possible to check for differential speededness and for patterns of omission, as well as to gain insight into the differential attractiveness of distractors for the two subgroups.

Item Classification

Before the analysis of the DIF data, Reading Comprehension and Sentence Completion items for Forms 1, 2, and 3 were classified by an SAT test development examiner according to (1) traditional content classification, (2) factors examined in the Wendler and Carlton (1987) paper, and (3) additional characteristics thought to be related to DIF for females.

As in the Wendler and Carlton exploratory study, all items were classified by the traditional SAT content categories listed below.

I. Reading Comprehension item type

- A. Category of passage
  1. Biological science
  2. Physical science
  3. Social studies
  4. Humanities
  5. Narrative
  6. Argumentative
- B. Length of passage
  1. Short (75–175 words)
  2. Medium (175–275 words)
  3. Long (375–475 words)
- C. Type of reading question
  1. Main idea
  2. Explicit statement
  3. Inference
  4. Application
  5. Author's logic
  6. Style or tone

II. Sentence Completion item type

- A. Content of sentence
  1. Aesthetics/philosophy
  2. World of practical affairs
  3. Science
  4. Human relationships
- B. Number of words blanked
  1. One blank
  2. Two blanks

Items for this study were also classified according to the following factors explored in the Wendler and Carlton 1987 paper:

1. 30-minute section in which the item appeared (1–6)
2. Total number of items in the section (V1–45, V2–40)
3. Presence of unpleasant, graphic, disastrous, or controversial references in Sentence Completion items
4. Degree of abstractness of Sentence Completion items (all items classified either abstract/general/philosophical or concrete/specific/pragmatic)

In addition to focusing on the above factors from the Wendler and Carlton study, we also classified items and passages according to the following features:

5. General content category (science versus nonscience) for both Reading Comprehension and Sentence Completion items
6. "Technical" versus "nontechnical" content in science reading passages
7. References to "true" versus "surface" science in Sentence Completion items

The above classifications were arrived at in the following manner. First, for the Reading Comprehension item type, the content of the Narrative, Argumentative, and Social studies passages was examined; each of these passages was combined with either the Science passages or the Humanities passages, or it was placed in an "indeterminate category."

Next, all those passages labeled Science were further classified as either "technical" or "nontechnical." These two categories should not be confused with "hard" versus "easy." Rather, the distinction was made between a passage that, for example, discussed in some detail a biological, chemical, or physical process or phe-

nomenon and a passage that, for example, discussed the life of a historic scientific figure or a general philosophy or view of science.

Finally, each reading question was carefully examined and categorized as to whether or not it involved or rewarded "outside knowledge" in the field or subject covered by the passage. Such questions, which were perhaps keyable without the test-taker's having read the passage, or which rewarded familiarity with the subject being discussed, tended to appear with Science passages, although some questions in one of the Humanities passages (in Form 1) were also classified this way.

Within the Sentence Completion item type, an attempt was made first to clarify the four traditional content categories—that is, to identify and separate those items that might readily fall into more than one category. After this step was completed, the test developer looked again at all items labeled Science and categorized them according to the degree of scientific content in the sentence. All Science Sentence Completion items were classified either "true science" or "surface science." As in the similar categorization of Reading Comprehension passages described above, such distinctions were not intended to be related to the difficulty of the item. Sentences about heavy snowfall, planetary features, or insect behavior were classified as true science. Items concerning the rancor of rival researchers, the industry of a laboratory technician, or the qualities of a particular scientific paper—that is, items whose context could easily be shifted to that of politics or art or economics—were classified as surface science.

A final content-relevant classification of all items identified those reading passages of special interest to female candidates, as well as any Sentence Completion items referring to a particular female or merely using a female pronoun (for example, "the doctor . . . she . . ."). This classification was used to determine the effects, if any, of such references on differential item functioning.

In addition to the aforementioned item content features, the following statistical factors were examined:

8. Difficulty of item and its relationship to DIF
9. Differential speededness on Reading Comprehension items related to the last passage in Verbal 2, which are the last items in that section
10. Differential rates of test-taker omission for discrepant items
11. Patterns of distractors for discrepant items

After analyzing data for Forms 1, 2, and 3, we attempted to verify salient factors on item response data from a fourth form. These last results are described in Confirmatory Analyses for Form 4 in the Results and Discussion section below.

RESULTS AND DISCUSSION

Because Reading Comprehension and Sentence Completion items were classified according to different characteristics, results are reported separately for the two item types. In addition, results are shown separately for each SAT form that was analyzed.

Distribution of DIF for the Four Verbal Item Types

Before we embark on a detailed description of the results for the Reading Comprehension and Sentence Completion items, it is important to convey a sense of the magnitude of differential item functioning (DIF) in the SAT-verbal test as a whole. Frequency distributions of standardized differences in proportion correct (STPDIF) for Forms 1, 2, and 3, respectively, broken down by item type, are shown in Tables 2, 3, and 4. Summary counts for different intervals of STPDIF are shown at the bottom of the table for each test form.

Inspection of these tables reveals that only a few items per form exhibit extreme negative or positive DIF (negative values indicate items that are more difficult for females; positive values indicate items that are easier for females). For Forms 1 and 2, the Vocabulary subscore (based on Antonyms and Analogies) contains more items with extreme values of negative DIF than does the Reading subscore. Relative to other forms, the distribution of STPDIF for Form 2 Reading Comprehension items is somewhat more truncated; fewer items exhibit extreme values of DIF.

For items constituting the Reading subscore, the focus of this investigation, the two most discrepant Reading Comprehension items have STPDIF values of −.11 (both appear in Form 1). The most discrepant Sentence Completion item has a STPDIF that is positive in value (.10) and appears in Form 2. The Reading Comprehension and Sentence Completion item types include relatively few differentially easy or difficult items in these three forms.

Reading Comprehension Items

Figures 1, 2, and 3 (pages 8 and 10) are pictorial displays of the STPDIF values associated with each Reading Comprehension item in the three forms initially studied. Because these figures are a fairly novel way of presenting data from DIF analyses, their components will be described briefly. Values of STPDIF (expressed as standardized differences in percentages correct) are indicated on the vertical axis (negative values indicate the item is more difficult for female examinees; positive values indicate the item is easier for female examinees). The locations of the Reading Comprehension items in the two SAT-

(continued on page 8)

Table 2. Distribution of Standardized P-Differences between Male and Female Examinees, by Item Type: Form 1

Vocabulary			*Reading*	
Antonyms	*Analogies*	*Value of STPDIF*	*Reading Comprehension*	*Sentence Completion*
		.15		
		.14		
		.13		
		.12		
		.11		
		.10		
		.09		
2		.08		1
1	2	.07		
		.06	1	
	1	.05		
		.04		2
1	1	.03	8	2
5	2	.02	1	4
5	3	.01	1	1
3	4	.00	6	2
3		−.01	1	1
1	2	−.02	1	
2		−.03	1	
1	1	−.04	2	1
	2	−.05		
	1	−.06		
		−.07		1
		−.08		
1		−.09	1	
		−.10		
		−.11	2	
		−.12		
		−.13		
		−.14		
	1	−.15		
		−.16		
		−.17		
		−.18		
		−.19		
		−.20		

	Antonyms	*Analogies*	*Reading Comprehension*	*Sentence Completion*	*Total*
STPDIF ≥ .10	0	0	0	0	0
.05 ≤ STPDIF < .10	3	3	1	1	8
−.05 ≤ STPDIF < .05	21	13	21	13	68
−.05 ≥ STPDIF > −.10	1	3	1	1	6
STPDIF ≤ −.10	0	1	2	0	3
Total	25	20	25	15	85

Table 3. Distribution of Standardized P-Differences between Male and Female Examinees, by Item Type: Form 2

Vocabulary			*Reading*	
Antonyms	*Analogies*	*Value of STPDIF*	*Reading Comprehension*	*Sentence Completion*
		.15		
		.14		
		.13		
		.12		
		.11		
	1	.10		1
		.09		
1		.08		
		.07		
1	1	.06	2	
		.05	2	
		.04	3	
3		.03	2	2
6	1	.02	3	1
1	3	.01	4	5
4	6	.00	4	3
3	1	−.01	1	1
1	3	−.02	1	1
		−.03		1
2		−.04	2	
2	1	−.05	1	
		−.06		
	1	−.07		
		−.08		
		−.09		
		−.10		
		−.11		
	1	−.12		
		−.13		
		−.14		
		−.15		
		−.16		
1		−.17		
		−.18		
		−.19		
	1	−.20		

	Antonyms	*Analogies*	*Reading Comprehension*	*Sentence Completion*	*Total*
STPDIF ≥ .10	0	1	0	1	2
.05 ≤ STPDIF < .10	2	1	4	0	7
−.05 < STPDIF < .05	20	14	20	14	68
−.05 ≥ STPDIF > −.10	2	2	1	0	5
STPDIF ≤ −.10	1	2	0	0	3
Total	25	20	25	15	85

Table 4. Distribution of Standardized P-Differences between Male and Female Examinees, by Item Type: Form 3

Vocabulary			*Reading*	
Antonyms	*Analogies*	*Value of STPDIF*	*Reading Comprehension*	*Sentence Completion*
		.15		
		.14		
		.13		
		.12		
1		.11		
		.10		
1		.09		
	2	.08	1	
		.07	1	
2	1	.06		
2		.05		
3		.04		1
2		.03	1	2
1	3	.02	2	1
3	1	.01	6	2
2	6	.00	4	4
	2	−.01	2	1
3	1	−.02	4	1
1	1	−.03	1	
1		−.04		1
	1	−.05	1	
		−.06	1	
2		−.07		
1	2	−.08	1	2
		−.09		
		−.10		
		−.11		
		−.12		
		−.13		
		−.14		
		−.15		
		−.16		
		−.17		
		−.18		
		−.19		
		−.20		

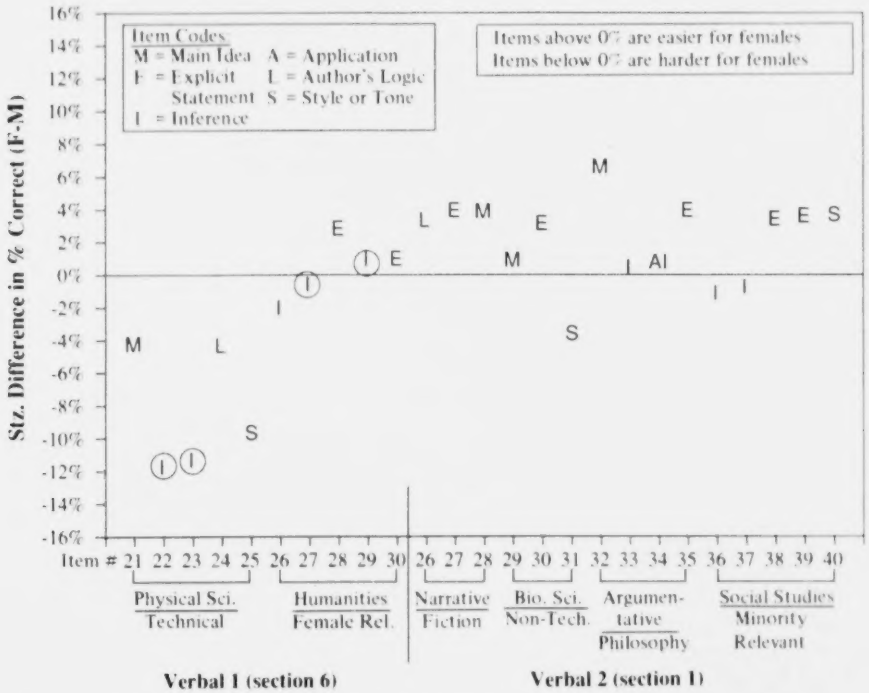
	Antonyms	*Analogies*	*Reading Comprehension*	*Sentence Completion*	*Total*
STPDIF ≥ .10	1	0	0	0	1
.05 ≤ STPDIF < .10	5	3	2	0	10
−.05 < STPDIF < .05	16	14	20	13	63
−.05 ≥ STPDIF > −.10	3	3	3	2	11
STPDIF ≤ −.10	0	0	0	0	0
Total	25	20	25	15	85

verbal sections are marked along the horizontal axis. The brackets beneath the item numbers are used to indicate sets of items appearing with each of the six separate reading passages. The next line on the horizontal axis indicates the subject matter of the reading passage as classified by SAT test developers for content specifications. The last line contains second-level descriptors of the passage; these descriptors were devised, as described earlier in the report, to provide additional ways of classifying the items that may be related to DIF for females. The letter or letters plotted in the body of the figure correspond to the item code used by SAT test developers for content specifications (for some items more than one code is appropriate, and in such cases two letters are used as a data point). The location of the letter or letters along the vertical axis indicates the STPDIF value for each Reading Comprehension item in the test. Circled items designate items identified as possibly rewarding examinees who have outside knowledge in the subject covered by the passage.

When Figure 1 is used as an example, it can be seen that the first Reading Comprehension item, 21, has a STPDIF value of −4 percent (−.04). The item has the following characteristics: it involves the main idea, or thesis, of the passage (*M*); it deals with a passage that has physical science content; the content is technical; and the item pertains to the first passage in the Verbal 1 section, which appeared in the last section (6) of the test.

Form 1 (Figure 1). Inspection of Figure 1 reveals that among items exhibiting negative DIF, the most discrepant items are 22, 23, and 25 in Verbal 1. Note that all of these items pertain to the same passage, which is a physical science passage and appears first in Verbal 1. While Form 1 also contains another science passage (involving biological science), that passage is not so technical. Note that for the nontechnical science passage, two of the three items are easier for females; one item (31) is slightly more difficult but clearly is not an outlier.

Turning our attention back to the three items (22, 23, and 25) exhibiting sizable negative DIF, we can see that two of the three require the examinee to make inferences from the material in the passage. While Form 1 contains other items classified "inference," those are not discrepant in terms of DIF. A further consideration of 22, 23, and 25 may reveal why these items are more difficult for females than for males. Recall that this passage deals with technical material: there may be an interaction between passage content and item type, so that additional background in chemistry, physics, biology, or other scientific fields could ad-



Note: Circles designate items identified as perhaps rewarding examinees having outside knowledge.

Figure 1. Standardized Percentage DIF Values for Males and Females on SAT Reading Comprehension Items: Form 1

vantage an examinee on the items requiring inferences from technical material. One might speculate that female examinees, relative to males of similarly developed verbal ability, bring less outside knowledge into the testing situation; if so, these items would seem to be easier for males, who may have had a broader background in science. While a warm-up effect is an unlikely explanation for the extreme negative DIF in these items (the passage appeared first in the last section of the test), there may also be a relationship between DIF and the location of a technical passage that is first in a 45-item section, preceded only by discrete items. Except for this first reading passage, the Reading Comprehension items in this form are generally easier for females.

Form 2 (Figure 2). It can be seen from Figure 2 (page 10) that Reading Comprehension items in Form 2 exhibit minimal DIF. Note that none of the passages in this test form contains technical science material. Moreover, one of the nontechnical science passages deals with content judged to be of special relevance to females. All the items pertaining to that passage are slightly easier for female examinees, suggesting that enhanced interest may have played a role in affecting item functioning for this passage.

Form 3 (Figure 3). Figure 3 reveals that for Form 3, four items related to the first passage in Verbal 1 are more difficult for females. Of these four items, however, only one (25) has a STPDIF value that is extreme. This item has several similarities to the discrepant items identified in Form 1; namely, it requires inference (or application) and is associated with a passage containing technical science material. As in Form 1, there is also the possibility that DIF is related to the first-place position of the technical passage in Verbal 1.

Form 3 has two other items exhibiting sizable DIF (27 and 28 in Verbal 2); these items are easier for female examinees. The corresponding Argumentative passage laments the apparent triumph of social conformity over individuality. There is nothing in particular about the passage or the items to explain why two items out of three are easier for female examinees than for male examinees. It is not sufficient to point out that the passage deals with nonscience content, because several other passages in this form deal with nonscience material and do not appear to favor female examinees.

Distractor Analyses for Reading Comprehension Items

Examples of Reading Comprehension items exhibiting extreme negative or positive DIF, along with plots of the conditional probability of choosing a particular response option (including the choice of omitting the item), are presented in Appendix A. (For security reasons, only items that are no longer in secure tests have been reproduced in Appendix A.)

The plots accompanying these items in Appendix A figures illustrate the empirical characteristic functions (ECFs) for each response option—that is, the percentage of the male and female subgroups selecting the option at each score level. Next to the ECF for each option is a corresponding difference plot for the same data. Symbols above the line indicate that females selected the option more often than males did. The correct option is denoted by an asterisk. For each response option, the STPDIF value, which is a weighted average of the differences across score levels, is displayed inside the difference plot. These plots provide information about differential item functioning at various levels along the ability continuum. Empirical characteristic functions for discrepant items were examined to determine whether certain distractors are differentially attractive to male or to female examinees. The presence of a high STPDIF value (absolute value at least .05) for a particular distractor could help to explain why an item is performing differentially for the two subgroups.

Table A.1 in Appendix A displays the first reading passage appearing in Verbal 1 of Form 1, along with the associated questions. The STPDIF value for each option is indicated in this table. Figures A.1 through A.3 present the empirical characteristic functions for all options for the three extreme items (22, 23, and 25) associated with this passage. The ECFs reveal several features of candidate response patterns by gender. On item 22 (Figure A.1), females tend to choose options *C* and *D* over the correct answer. The distractor pattern for item 23 (Figure A.2) is more clear-cut: option *B* is selected more often by females than by males, and the difference in responding is most pronounced in the middle portion of the scaled score range.

No particular content-related reasons for the differential option selection were identified for items 22 and 23, and there were no discernible omit-rate differences. However, the distractor analysis for item 25 (Figure A.3) lends support to the hypothesis concerning the technical nature of the science passage. For item 25, option *E* is selected more often by females, particularly below the scaled score of 500; option *A* is selected, but to a much lesser extent, by females from the middle to the top of the score range. This question asks the examinee to identify the intended audience of the passage (Table A.1). The correct response is keyed as option *D* for "general readers," but females select option *E* (especially) and option *A* more often than do males, indicating female perception of the intended audience as "researchers" or "advanced students." This perception of the "complexity" of the passage is perhaps related to differential background, interest, or self-perceived capability with technical science material.

Table A.2 in Appendix A contains the first reading

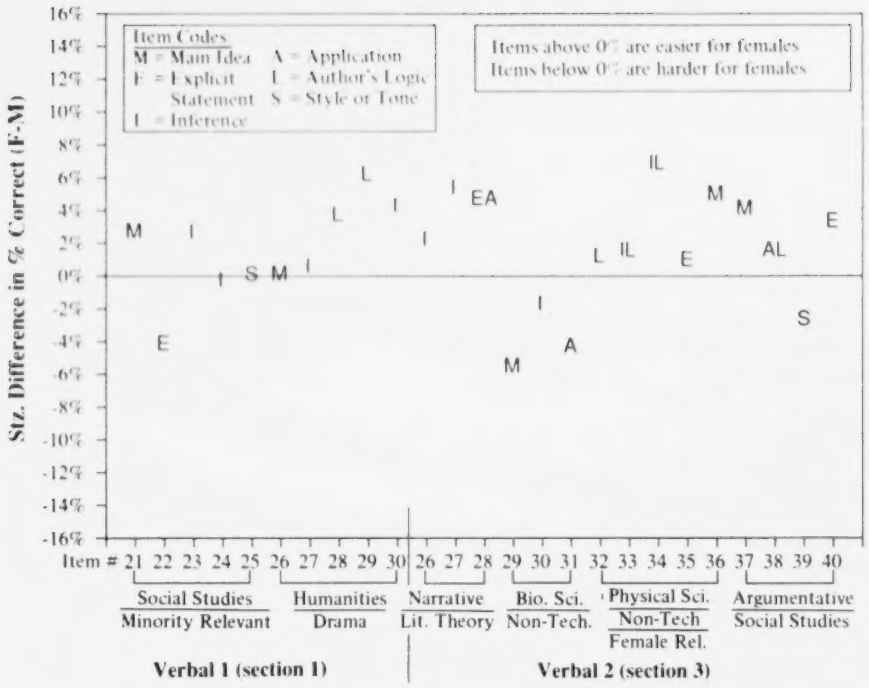
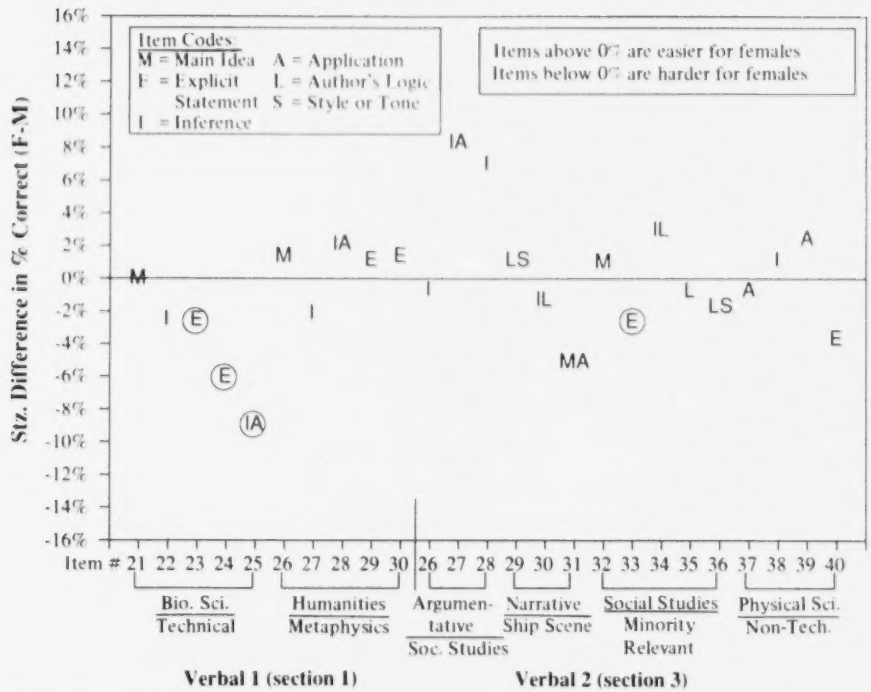

Item Codes
M = Main Idea
E = Explicit Statement
I = Inference
A = Application
L = Author's Logic
S = Style or Tone
Items above 0% are easier for females
Items below 0% are harder for females
Std. Difference in % Correct (F-M)
Item # 21 22 23 24 25 26 27 28 29 30 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40
Social Studies / Minority Relevant
Humanities / Drama
Narrative / Lit. Theory
Bio. Sci. / Non-Tech.
Physical Sci. / Non-Tech / Female Rel.
Argumentative / Social Studies
Verbal 1 (section 1)
Verbal 2 (section 3)

Figure 2. Standardized Percentage DIF Values for Males and Females on SAT Reading Comprehension Items: Form 2


Item Codes
M = Main Idea
E = Explicit Statement
I = Inference
A = Application
L = Author's Logic
S = Style or Tone
Items above 0% are easier for females
Items below 0% are harder for females
Std. Difference in % Correct (F-M)
Item # 21 22 23 24 25 26 27 28 29 30 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40
Bio. Sci. / Technical
Humanities / Metaphysics
Argumen-tative / Soc. Studies
Narrative / Ship Scene
Social Studies / Minority Relevant
Physical Sci. / Non-Tech.
Verbal 1 (section 1)
Verbal 2 (section 3)

Note: Circles designate items identified as perhaps rewarding examinees having outside knowledge.

Figure 3. Standardized Percentage DIF Values for Males and Females on SAT Reading Comprehension Items: Form 3

passage appearing in Verbal 1 of Form 3. Figure A.4 presents the ECFs for the most extreme item (25) associated with this passage. The plots indicate that no particular distractor is selected differentially. This item is more difficult for female examinees, particularly for examinees with scaled scores between 400 and 500. Note that the omit pattern is slightly different for males and females: a larger proportion of female examinees elect to omit this item.

Table A.3 in Appendix A has the first reading passage and associated questions appearing in Verbal 2 of Form 3. The ECFs in Figures A.5 and A.6 (pp. 38–41) reveal that most of the differential functioning for the two extreme items (27 and 28) is among lower-ability examinees; in this portion of the ability range these two items are easier for female examinees.

Other Findings for Reading Comprehension Items. All the Reading Comprehension items exhibiting sizable DIF were inspected for the presence of other factors. No consistent trends were observed when items or passages were classified according to the length of the passage (see Table 5, p. 13). Length of passage was confounded in technical science material, however, in that both technical passages in these three forms were also long. With respect to subject matter, female interest, or item type as an explanation of DIF, Figures 1 to 3 clearly demonstrate that such factors are indistinguishable individually as explanatory characteristics. Item difficulty was also examined as a potential explanatory factor. Within the three initial forms studied, item difficulty and DIF do not appear to be related (see Figures 4 through 6).

The clearest trend in the three forms is a distinction between technical and nontechnical science content, with an interesting contrast between Form 1 and Form 3. While Form 1 has a technical physical science passage and a nontechnical biological science passage, Form 3 reflects the reverse configuration, with a technical biological science passage and a nontechnical physical science passage. Note that the substantial negative STPDIF values for the items classified "inference" correspond to the technical passage, independent of the science discipline.

The effect of distinguishing among passages on the basis of technical versus nontechnical science content is best seen in Table 6 (p. 13), where STPDIF values are broken down by science content for the three forms. The left-hand third of Table 6 shows that virtually all the items pertaining to technical science passages are more difficult for female examinees. Furthermore, all the substantial negative STPDIF values are associated with items involving technical science passages.

It is important to emphasize that technical science content is *not* synonymous with item difficulty. In Form 1, three of the five items associated with the technical science passage are of medium difficulty and two are

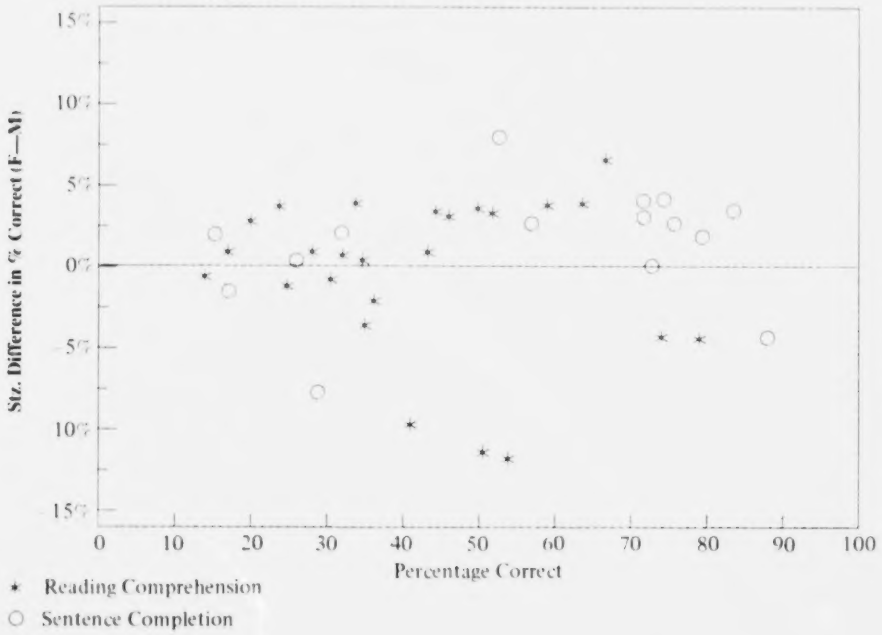


Figure 4. Standardized Percentage DIF Values for Males and Females, by Item Difficulty: Form 1

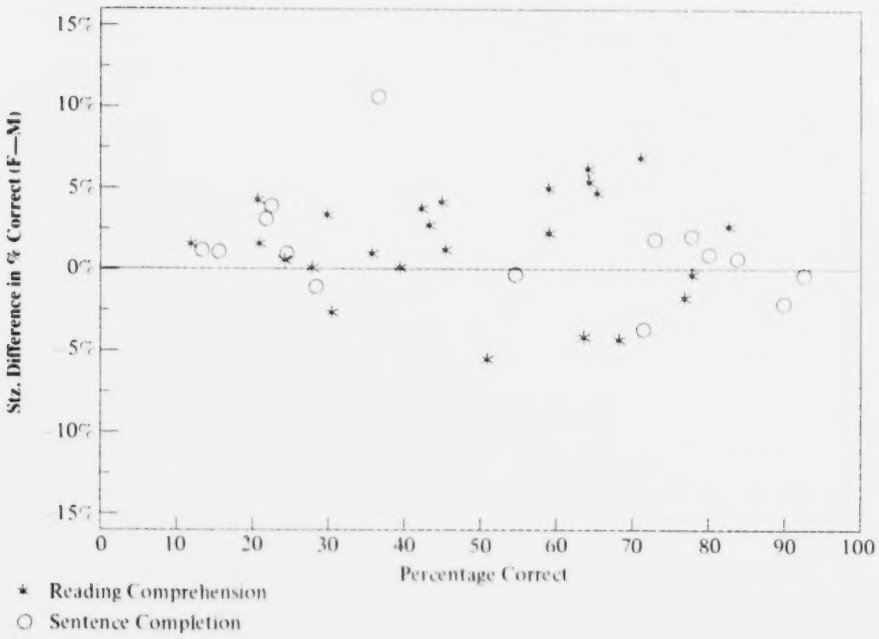


Figure 5. Standardized Percentage DIF Values for Males and Females, by Item Difficulty: Form 2

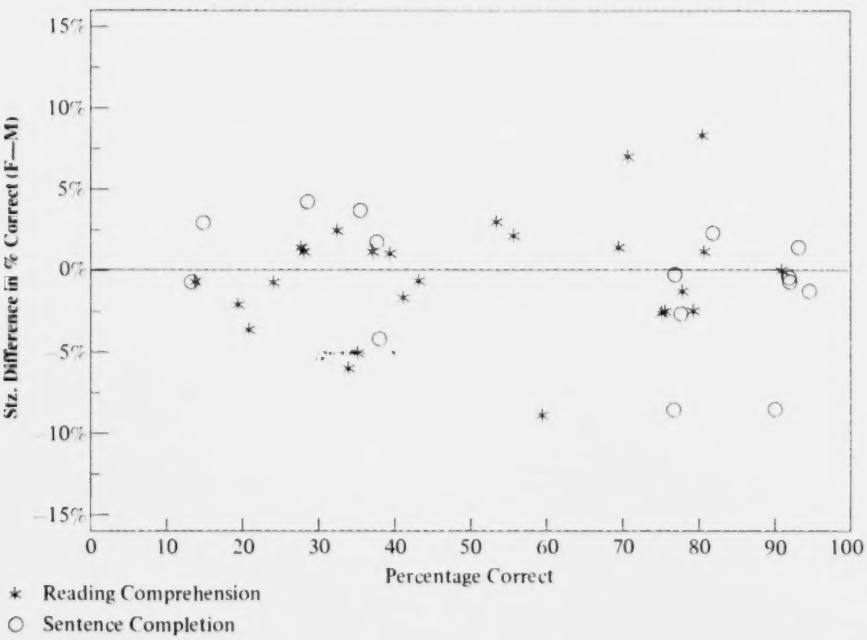


Figure 6. Standardized Percentage DIF Values for Males and Females, by Item Difficulty: Form 3

Table 5. Distribution of Standardized P-Differences between Male and Female Examinees, by Length of Reading Passage

	Long Passage	*Medium Passage*
STPDIF ≥ .10	0	0
.05 ≤ STPDIF < .10	3	3
−.05 < STPDIF < .05	37	25
−.05 ≥ STPDIF > −.10	3[a]	2
STPDIF ≤ −.10	2[a]	0
Total	45	30

Note: Frequencies are based on Reading Comprehension items for Forms 1, 2, and 3, for a total of 75 items across the three forms.

a. All these items are associated with "technical" science reading passages.

Table 6. Pictorial Diagram of Standardized P-Differences between Male and Female Examinees on Reading Comprehension Items, by Science Content

Technical				*Nontechnical*				*Nonscience*		
Form 1	*Form 2*	*Form 3*	*Value of STPDIF*	*Form 1*	*Form 2*	*Form 3*	*Value of STPDIF*	*Form 1*	*Form 2*	*Form 3*
			.11				.11			
			.10				.10			
			.09				.09			
			.08				.08			R3
			.07		R5		.07			R3
			.06				.06	R5	R2	
			.05		R5		.05		R3	
			.04				.04		R2 R3 R6	
			.03	R4			.03	R3 R3 R3 R5 R6 R6 R6	R2 R6	R5
			.02			R6	.02	R2	R1 R1 R3	R2
			.01		R5 R5 R5	R6	.01	R2	R6	R2 R2 R2 R4 R5
		R1	.00	R4		R6	.00	R2 R2 R5 R5 R6	R1 R1 R2 R2	R3 R5
			−.01		R4		−.01	R6		R4 R5
		R1 R1	−.02				−.02	R2	R6	R2 R5
			−.03	R4		R6	−.03			
R1 R1			−.04		R4		−.04		R1	
			−.05		R4		−.05			R4
		R1	−.06				−.06			
			−.07				−.07			
		R1	−.08				−.08			
R1			−.09				−.09			
			−.10				−.10			
R1 R1			−.11				−.11			

Note: R followed by a number refers to reading passage number (Verbal 1 contains passages R1 and R2; Verbal 2 contains passages R3 through R6).

easy. In Form 3, three of the items associated with the technical science passage are easy and two are of medium difficulty. Thus, none of the items associated with these technical science passages is very difficult; in fact, the items associated with nontechnical science passages are slightly more difficult overall. Table 7, which presents item difficulty distributions in equated delta units, broken down by technical versus nontechnical science content, documents that the technical science items are, in fact, slightly easier.

Speededness. Certain Reading Comprehension items were examined for differential speededness. In particular, the items associated with the last passage in Verbal 2 are located in the last three, four, or five positions in that section, depending on the form. Findings from distractor analyses indicate that across the three test forms initially studied, the proportion of female examinees not reaching the last set of items in Verbal 2 is slightly lower than the proportion of male examinees with the same ability. However, these standardized differences are very small (ranging between −.01 and −.04). Correcting differential item function statistics for the effect of speededness (computing STPDIF using only those examinees who reached the item) had virtually no effect on the magnitude or the direction of the STPDIF values.

Sentence Completion Items

The results of the analyses for Sentence Completion items are shown in Figures 7 to 9. As was done for the Reading Comprehension items, the layout of these figures will be described briefly. Values of STPDIF are indicated on the vertical axis (negative values indicate the item is more difficult for female examinees; positive values indicate the item is easier for female examinees). The locations of the Sentence Completion items in the two SAT-verbal sections are marked along the horizontal axis. For each item the letter or letters plotted in the body of the figure correspond to the item code used by SAT test developers for content specifications (for some items more than one code is appropriate). The location of the letter or letters along the vertical axis indicates the STPDIF value for each Sentence Completion item in the test. The labels below certain item numbers on the horizontal axis designate whether the science reference is to true science or to surface science (see earlier discussion under Procedures). When Figure 7 is used as an example, it can be seen that item 16 has a STPDIF value of 2 percent (.02). This item appeared in Verbal 1, which was given as section 6 in the operational test. The content of this item deals with aesthetics/philosophical material (*A*).

Table 7. Difficulty Distributions in Equated Delta Units, by Form and Science Content: Reading Comprehension Items

Technical Science			*Equated*	*Nontechnical Science*		
Form 1	*Form 2*	*Form 3*	*Delta*	*Form 1*	*Form 2*	*Form 3*
			19			
			18			
			17			
			16		1	1
			15			
	None		14			2
		1	13	1		1
1			12	1	1	
2		1	11	1	2	
			10		1	
		1	9		1	
1		1	8		2	
1		1	7			
			6			
			5			
5	0	5	*N* items	3	8	4

Form 1 (Figure 7). In this form, two items (18 in Verbal 1 and 14 in Verbal 2) are extreme enough to be viewed as performing differentially for female examinees. On the basis of item content alone, consistent trends are not observed. Although one item dealing with human relationships exhibits extreme DIF in the positive direction, other items with similar content do not follow this trend. The item exhibiting extreme DIF in the negative direction has science content. Note, however, that of the four items with science content, two are easier for females and two are more difficult for females. In terms of a section effect, the two items exhibiting negative DIF are located in Verbal 2, and the two items exhibiting positive DIF are located in Verbal 1. However, section is confounded with another factor. The items that are more difficult for females are true science items, and both appear in Verbal 2, while the items that are easier for females are surface science items, and both appear in Verbal 1.

Form 2 (Figure 8). With the exception of one item (14) in Verbal 2, this form appears to be void of differentially easy or difficult Sentence Completion items. The one item that is easier for females is classified "human relationships," but the other items in this classification do not exhibit similar DIF. Hence, it cannot be concluded that items whose content is related to human relationships necessarily favor female examinees. Note, for instance, that of the "pure" human relationship items (coded *H*, 32, 33, and 14), all are easier for females, albeit slightly, while items with world of practical affairs/human relationship content (WH) have STPDIF values that are less positive.

Form 3 (Figure 9). In this form, there are two discrepant Sentence Completion items (18 in Verbal 1 and 13

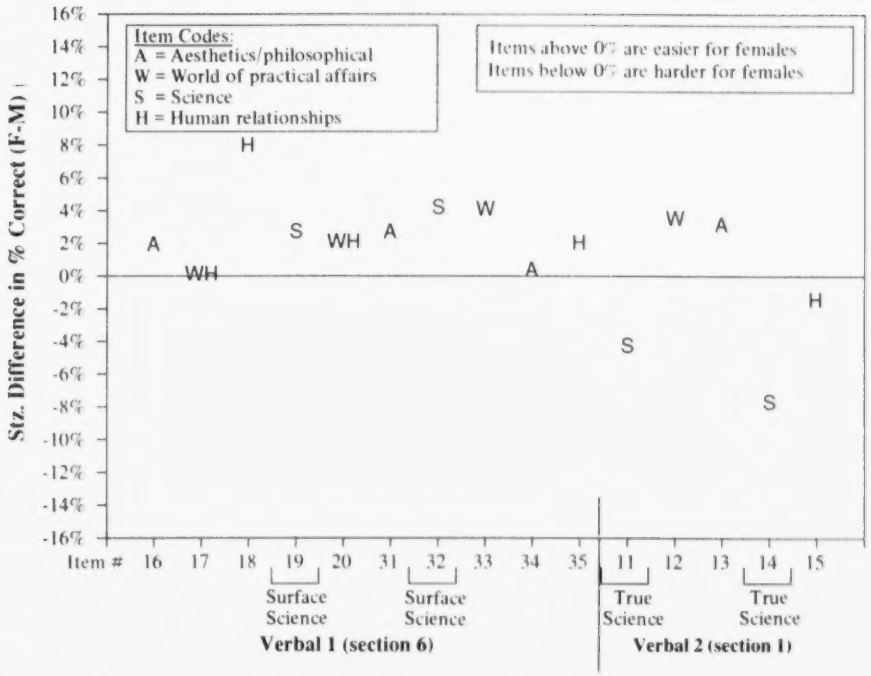

Item Codes:
A = Aesthetics/philosophical
W = World of practical affairs
S = Science
H = Human relationships
Items above 0% are easier for females
Items below 0% are harder for females
Stz. Difference in % Correct (F-M)
Item # 16 17 18 19 20 31 32 33 34 35 11 12 13 14 15
Surface Science
Surface Science
True Science
True Science
Verbal 1 (section 6)
Verbal 2 (section 1)

Figure 7. Standardized Percentage DIF Values for Males and Females on SAT Sentence Completion Items: Form 1

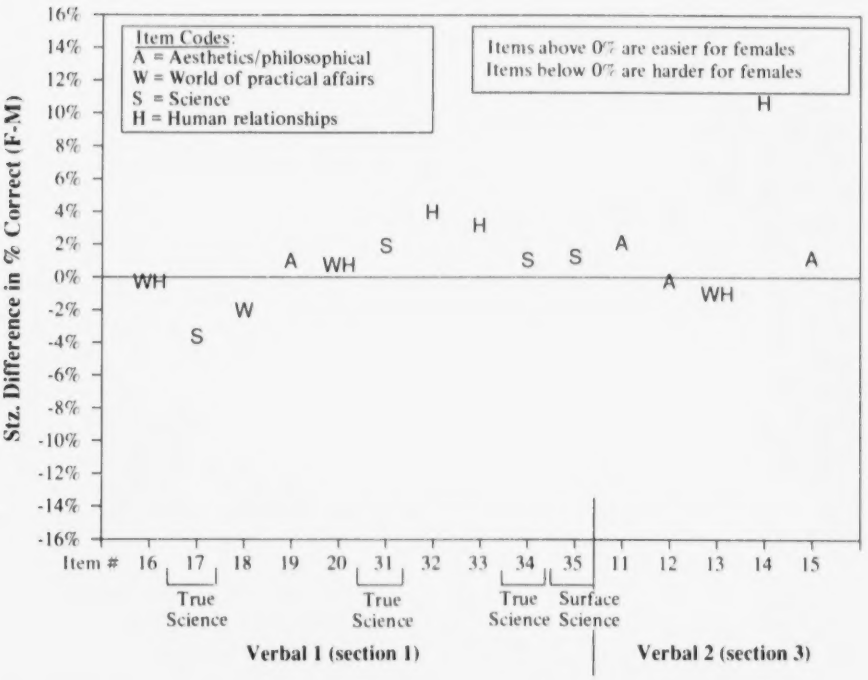

Item Codes:
A = Aesthetics/philosophical
W = World of practical affairs
S = Science
H = Human relationships
Items above 0% are easier for females
Items below 0% are harder for females
Stz. Difference in % Correct (F-M)
Item # 16 17 18 19 20 31 32 33 34 35 11 12 13 14 15
True Science
True Science
True Science
Surface Science
Verbal 1 (section 1)
Verbal 2 (section 3)

Figure 8. Standardized Percentage DIF Values for Males and Females on SAT Sentence Completion Items: Form 2

in Verbal 2), and both have true science content. Of the five science-content items, three have STPDIF values close to zero. Of these three items, however, two involve only surface references to science. Of the two science-content items that are clearly more difficult for female examinees, both involve true science references.

Distractor Analyses for Sentence Completion Items

Appendix B contains item text and empirical characteristic function (ECF) plots for two Sentence Completion items exhibiting extreme positive or negative DIF in the three forms studied initially. The actual items are reproduced in Table B.1 in Appendix B (only items no longer secure have been reproduced), with STPDIF values for each response option.

An example of an extreme Sentence Completion item in Form 1 is item 14 in Verbal 2. This item is more difficult for female examinees throughout most of the ability range (see ECFs for this item in Appendix B, Figure B.1). The distractor analysis is not helpful in identifying the source of differential functioning for this item because no one distractor is selected.

Form 3 provides an example of a Sentence Completion item exhibiting extreme DIF: 18 in Verbal 1 (STPDIF = −.0847). The ECF plots in Figure B.2 for item 18 indicate that for males and females with scaled scores below 600, the item functions differently. There is a slight tendency for low-ability females to choose option *B* instead of the correct answer. Note that this is a very easy item (equated delta 6.8); consequently, most examinees in the upper portion of the ability range answer the item correctly, regardless of gender. Among lower-ability examinees, however, the difference in performance by males and females is substantial.

Other Findings for Sentence Completion Items. All the Sentence Completion items were classified in terms of difficulty, a characteristic that was not found to be consistently related to DIF (see Figures 4, 5, and 6). Also, the distributions of equated deltas for items broken down by true versus surface science content are similar (see Table 8). In addition, analyses of STPDIF values in terms of the number of blanks in the item, negative tone (that is, the presence of unpleasant, graphic, or controversial material) in the item, and the degree of abstractness in the item failed to uncover any salient or systematic findings (Table 9). Item position and gender reference in the item are similarly unrelated to the incidence of DIF.

The plots in Figures 7 to 9 provide clear evidence that content codes are not related in a systematic way to DIF. For the Sentence Completion items, the most prominent explanatory characteristic appears to involve the distinction between items having true references to

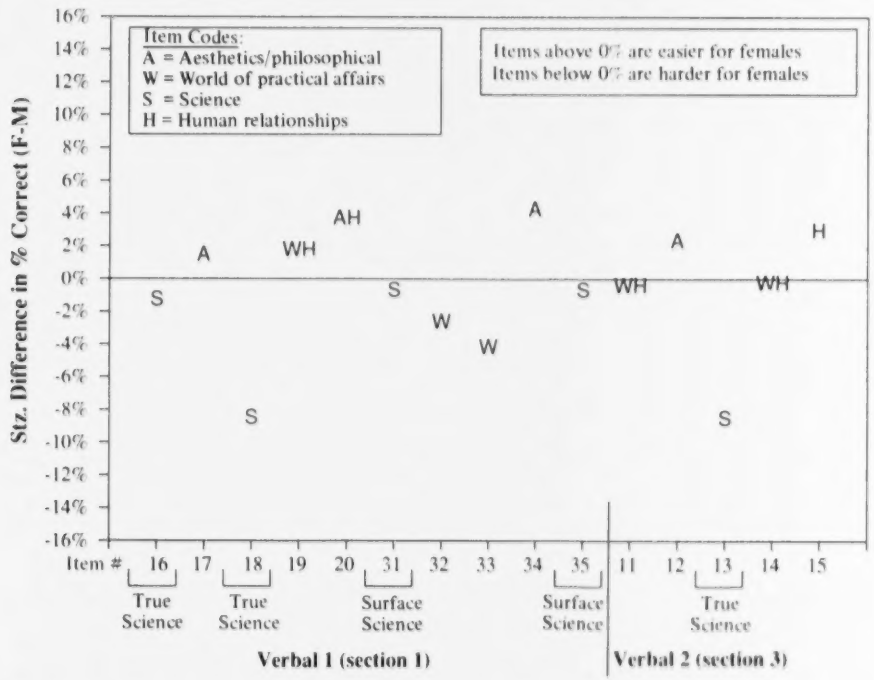


Figure 9. Standardized Percentage DIF Values for Males and Females on SAT Sentence Completion Items: Form 3

Table 8. Difficulty Distributions in Equated Delta Units, by Form and Science Content: Sentence Completion Items

True Science			*Equated*	*Surface Science*		
Form 1	*Form 2*	*Form 3*	*Delta*	*Form 1*	*Form 2*	*Form 3*
			19			
			18			
			17			
			16		1	1
			15			
	1		14			
1			13			
			12			
			11			
			10	1		
		1	9			
	2		8	1		
			7			
		2	6			1
1			5			
2	3	3	*N* items	2	1	2

science and those having surface references to science. In general, Sentence Completion items with true references to science are more difficult for female examinees. This trend is shown in Table 10, where it can be seen that all items with true science reference have STPDIF values of .01 and below, with several below −.04. On the other hand, the distributions of STPDIF for items with surface science reference are centered near zero and are similar to the distributions for nonscience items. In other words, the Sentence Completion items with true science content seem to behave slightly differently from all the other Sentence Completion items. Note also that the only items with STPDIF values greater than .05 are nonscience items.

Note that two out of three Sentence Completion items exhibiting extreme negative DIF are located in Verbal 2. However, this characteristic is confounded by the presence of true science content. The data for Form 3 argue against the distinction between sections as an explanation: while both of the negatively discrepant items have true science content, one of the items appears in Verbal 1, the other in Verbal 2.

Summary of Findings for Exploratory Analyses of Forms 1, 2, and 3

The findings described thus far point to several tentative conclusions concerning differences in item functioning for male and female examinees who have been matched on ability. *The primary finding is that minimal DIF was observed in the three forms.* Of the few extreme items, the main factor that appears to be related to DIF for females on Reading Comprehension items is the distinction between technical and nontechnical reading passage material. Items associated with passages containing technical science material are generally more difficult for female examinees. The main factor that appears to be related to DIF for females on Sentence Completion items is the distinction between nonscience or surface science reference and true science reference. Items containing true science references tend to be more difficult for females. Given the apparent salience of the true science/surface science dichotomy for Sentence Completion items and the technical science/nontechnical science dichotomy for Reading Comprehension items, an attempt was made to verify the explanatory power of these factors on data from another, fourth test form. These findings are described below.

Confirmatory Analyses for Form 4

Based on findings from previous analyses, items and passages in Form 4 were classified according to the following features:

1. Science content versus nonscience content for both Reading Comprehension and Sentence Completion items
2. Technical versus nontechnical science content in reading passages
3. References to true science versus surface science in Sentence Completion items

As was done for Forms 1, 2, and 3, each reading question was also carefully examined and categorized as to whether the question may involve or reward outside knowledge in the field or subject covered by the passage.

Frequency distributions of standardized differences in the proportion correct (STPDIF) for Form 4, broken down by item type, are shown in Table 11 (p. 19). Summary counts for different intervals of STPDIF are shown at the bottom of the table. Overall, the Reading subscore contains fewer items with extreme values of negative DIF than does the Vocabulary score. However, the most extreme items in the test are a Reading Comprehension item (STPDIF value of −.15) and a Sentence Completion item (STPDIF value of .15). With the exception of the Analogy item type, where nearly half the items have STPDIF values below −.05, the items in this form exhibit minimal DIF for males and females: most of the STPDIF values are distributed between −.05 and .05.

Reading Comprehension Items. Among the items exhibiting extreme negative DIF, only item 29 in Verbal 1 is an outlier (see Figure 10). While item 29 is a discrepant item, with a STPDIF value of −.15 (−15 percent), the other items pertaining to the first passage are not behaving similarly. Examination of the content of the passage reveals that it deals with military history. Although mili-

(continued on page 21)

Table 9. Distribution of Standardized P-Differences between Male and Female Examinees on Sentence Completion Items: Factors Identified in Previous Research

	Items with Negative Tone	*N* Blanks		*Abstract/Pragmatic*	
		1	*2*		
STPDIF ≥ .10	0	1	0	1	0
.05 ≤ STPDIF < .10	0	1	0	1	0
−.05 < STPDIF < .05	3	20	20	34	6
−.05 ≥ STPDIF > −.10	2	1	2	2	1
STPDIF ≤ −.10	0	0	0	0	0
Total	5	23	22	38	7

Note: Frequencies are based on the Sentence Completion items in Forms 1, 2, and 3, for a total of 45 items across the three forms.

Table 10. Pictorial Diagram of Standardized P-Differences between Male and Female Examinees on Sentence Completion Items, by Science Content

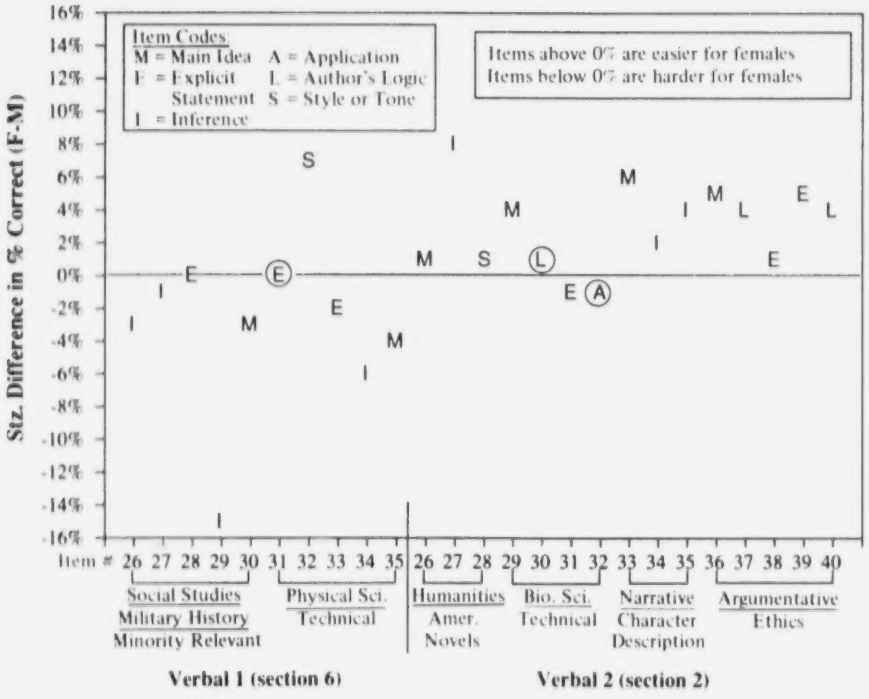
True Science			*Value of STPDIF*	*Surface Science*			*Value of STPDIF*	*Nonscience*		
Form 1	*Form 2*	*Form 3*		*Form 1*	*Form 2*	*Form 3*		*Form 1*	*Form 2*	*Form 3*
			.10				.10		V2	
			.09				.09			
			.08				.08	V1		
			.07				.07			
			.06				.06			
			.05				.05			
			.04	V1			.04	V1		V1
			.03				.03	V2 V2	V1 V1	V1 V2
			.02	V1			.02	V1 V1 V1	V2	V2
	V1 V1		.01				.01	V1	V1 V2	V1 V1
			.00		V1	V1 V1	.00	V1 V1	V1 V1 V2	V2 V2
		V1	−.01				−.01	V2	V2	
			−.02				−.02		V1	V1
	V1		−.03				−.03			
V2			−.04				−.04			V1
			−.05				−.05			
			−.06				−.06			
V2			−.07				−.07			
		V2 V1	−.08				−.08			
			−.09				−.09			
			−.10				−.10			

Note: V1 and V2 refer to the 45-item Verbal 1 section (Sentence Completion items are located in positions 16–20 and 31–35) and the 40-item Verbal 2 section (Sentence Completion items are located in positions 11–15).

Table 11. Distribution of Standardized P-Differences between Male and Female Examinees, by Item Type: Form 4

Vocabulary		*Value of STPDIF*	*Reading*	
Antonyms	*Analogies*		*Reading Comprehension*	*Sentence Completion*
		.15		1
		.14		
1		.13		
		.12		
		.11		
		.10		
		.09		
	1	.08	1	
1		.07	1	1
1	1	.06	1	1
1		.05	2	
2		.04	4	2
2	1	.03		
3	2	.02	1	3
1	2	.01	4	2
3	1	.00	2	1
5	2	−.01	3	1
1	1	−.02	1	1
		−.03	2	
1		−.04	1	
	2	−.05		1
	3	−.06	1	1
1	1	−.07		
	2	−.08		
1	1	−.09		
		−.10		
		−.11		
1		−.12		
		−.13		
		−.14		
		−.15	1	
		−.16		
		−.17		
		−.18		
		−.19		
		−.20		

	Antonyms	*Analogies*	*Reading Comprehension*	*Sentence Completion*	*Total*
STPDIF ≥ .10	1	0	0	1	2
.05 ≤ STPDIF < .10	3	2	5	2	12
−.05 < STPDIF < .05	18	9	18	10	55
−.05 ≥ STPDIF > −.10	2	9	1	2	14
STPDIF ≤ −.10	1	0	1	0	2
Total	25	20	25	15	85



Note. Circles designate items identified as perhaps rewarding examinees having outside knowledge.

Figure 10. Standardized Percentage DIF Values for Males and Females on SAT Reading Comprehension Items: Form 4

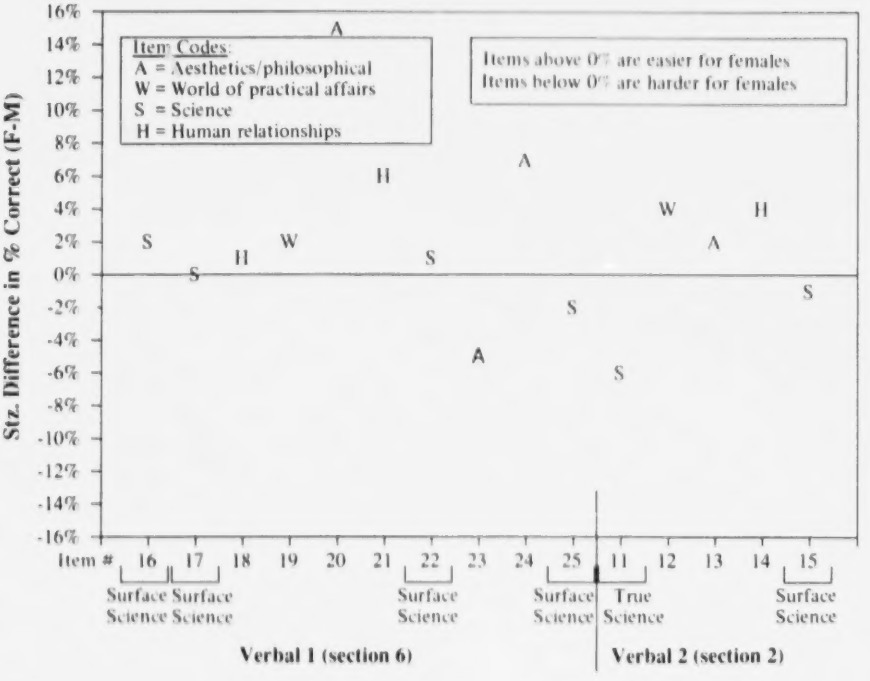

Item Codes:
A = Aesthetics/philosophical
W = World of practical affairs
S = Science
H = Human relationships
Items above 0% are easier for females
Items below 0% are harder for females
Stz. Difference in % Correct (F-M)
Item # 16 17 18 19 20 21 22 23 24 25 11 12 13 14 15
Surface Science
Surface Science
Surface Science
Surface Science
True Science
Surface Science
Verbal 1 (section 6)
Verbal 2 (section 2)

Figure 11. Standardized Percentage DIF Values for Males and Females on SAT Sentence Completion Items: Form 4

tary history is not hypothesized in this study to be a contributing factor to DIF, it is possibly a curricular area in which males study more frequently than females do. In terms of DIF in the positive direction, the most extreme item, 27 in Verbal 2, relates to a passage about American novels.

As can be seen in Table 12, results pertaining to the hypothesis that technical science passages are more difficult for females than for males did not conform to expectations. Form 4 has two passages containing technical science material: a physical science passage in Verbal 1 and a biological science passage in Verbal 2. For the physical science passage, one item (32) is easier for females; this particular item deals with the style or tone of the passage and not with the technical content. Item 34, on the other hand, requires the examinee to make an inference about the technical material presented in the passage; this item is somewhat more difficult for female examinees. Note that the items in Forms 1, 2, and 3 that were more difficult for females tended to be inference items dealing with technical science material.

Contrary to expectation, males and females do not appear to perform differentially on items corresponding to the biological science passage. One possible explanation for the finding is that this particular science passage does not have any inference items associated with it—that is, items where a correct answer may require particularly careful reading of the details and the substance of the technical material. Another possible explanation for failure to replicate the results observed for Forms 1, 2, and 3 is the location of the biological science passage. Note that for the forms described earlier, the technical science passages appeared in Verbal 1. Form 4 provides the first opportunity to observe a technical science passage in Verbal 2. This form also includes the first medium-length technical science passage rather than a long passage. To untangle the effects of passage location, length, and technical content, there is a need for experimental research in which passage location and other item and passage features can be carefully controlled.

The items associated with the last reading passage in Verbal 2 were examined for differential speededness. Similar to the pattern observed for Forms 1, 2, and 3, the proportion of female examinees not reaching this last set of Reading Comprehension items is slightly lower than the proportion of male examinees with similar ability who do not reach the last set (standardized differences for the "not-reaching" response range between −.03 and −.04). This indicates that matched groups of males and females behave slightly differently with respect to speededness; females reach more items than do males. However, correcting the DIF statistics for speededness had a negligible effect on the size or direction of the STPDIF values (the value for one item changed from .04 to .05).

Table 12. Pictorial Diagram of Standardized P-Differences between Male and Female Examinees on Reading Comprehension Items, by Science Content: Form 4

Technical		*Nontechnical*		*Nonscience*
	.10		.10	
	.09		.09	
	.08		.08	R3
R2	.07		.07	
	.06	None	.06	R5
	.05		.05	R6 R6
R4	.04		.04	R5 R6 R6
	.03		.03	
	.02		.02	R5
R4	.01		.01	R3 R3 R6
R2	.00		.00	R1
R4 R4	−.01		−.01	R1
R2	−.02		−.02	
	−.03		−.03	R1 R1
R2	−.04		−.04	
	−.05		−.05	
R2	−.06		−.06	
	−.07		−.07	
	−.08		−.08	
	−.09		−.09	
	−.10		−.10	
	−.11		−.11	
	−.12		−.12	
	−.13		−.13	
	−.14		−.14	
	−.15		−.15	R1

Note: R followed by a number refers to reading passage number (Verbal 1 contains passages R1 and R2; Verbal 2 contains passages R3 through R6).

Distractor Analyses for Reading Comprehension Items. The first reading passage and set of questions in Verbal 1 of Form 4 is reproduced in Appendix A in Table A.4, along with the empirical characteristic functions in Figure A.7 for the one extreme item (29) associated with the passage. Item 29 is considerably more difficult for female examinees. The plot of conditional responding reveals that this item is more difficult for female examinees throughout most of the ability range. Option *B* is more attractive to female examinees, particularly for examinees in the middle portion of the ability continuum.

Another reading passage and set of questions, in Verbal 2 of Form 4, is reproduced in Table A.5 (p. 29) in Appendix A, along with the ECFs in Figure A.8 (p. 45) for the one extreme item (27) associated with the passage. Item 27 is easier for female examinees throughout most of the scaled score range. At the top of the score range there is a tendency for more male examinees to incorrectly choose option *D*.

Sentence Completion Items. As can be seen in Figure 11, only one Sentence Completion item in this form (20 in Verbal 1) is clearly easier for female examinees. The

item deals with aesthetics/philosophical material and has a STPDIF value of .15 (15 percent).

Findings for the science-content items in Form 4 appear to substantiate the results observed in the previous forms, where items with true science content tended to be more difficult for female examinees. Of the six items in Figure 11 with science-related content, five were classified as surface science, and one (11) was classified as true science. As expected, item 11 is more difficult for females (STPDIF = −.06). This pattern of results is illustrated in Table 13.

Distractor Analyses for Sentence Completion Items. In Appendix B the ECF plots associated with item 20 (Verbal 1) are shown in Figure B.3 (p. 52), and the item is reproduced in Table B.1. Among examinees in the low-to-middle portion of the ability continuum (between scaled scores 300 and 600), the item is easier for females. The ECF for option *A* reveals that this particular distractor functions differentially for examinees with scaled scores below 600: relative to females with similar ability, a higher proportion of males select this option.

Table 13. Pictorial Diagram of Standardized P-Differences between Male and Female Examinees on Sentence Completion Items, by Science Content: Form 4

True Science		*Surface Science*		*Nonscience*
	.15		.15	V1
	.14		.14	
	.13		.13	
	.12		.12	
	.11		.11	
	.10		.10	
	.09		.09	
	.08		.08	
	.07		.07	V1
	.06		.06	V1
	.05		.05	
	.04		.04	V2 V2
	.03		.03	
	.02	V1	.02	V1 V2
	.01	V1	.01	V1
	.00	V1	.00	
	−.01	V2	−.01	
	−.02	V1	−.02	
	−.03		−.03	
	−.04		−.04	
	−.05		−.05	V1
V2	−.06		−.06	
	−.07		−.07	
	−.08		−.08	
	−.09		−.09	
	−.10		−.10	

Note: V1 and V2 refer to the 45-item Verbal 1 section (Sentence Completion items are located in positions 16–25) and the 40-item Verbal 2 section (Sentence Completion items are located in positions 11–15).

The ECF plots for item 11 (Verbal 2) are displayed in Figure B.4, and the item is reproduced in Table B.1. Among low-ability examinees, this true science item is more difficult for females. However, for examinees with scaled scores above 500, the item is equally difficult for males and for females. In terms of differential functioning for a particular distractor, the ECF for option *A* shows that low-ability males select option *A* in greater proportion than do females of similar ability.

CONCLUSIONS

This investigation was carried out to study whether certain SAT Reading Comprehension and Sentence Completion items exhibit differential item functioning (DIF) for female examinees relative to male examinees. Specifically, the goal of the research was to determine whether DIF can be attributed to particular factors concerning the items. The findings described in this report have led to several conclusions about differences in item functioning for the two subgroups. It is important, however, to stress that the conclusions are based on limited occurrences of substantial DIF. Across the four test forms studied in this investigation, values of STPDIF ranged between −.05 and .05 for most items in the Reading subscore. Among the few items exhibiting substantial DIF, two trends related to science content were noted.

Reading Comprehension Items. An important factor that appears to be connected to DIF on Reading Comprehension items is the distinction between technical and nontechnical reading passage material. Items associated with passages containing technical (as opposed to historical or philosophical) science material are generally more difficult for female examinees. This trend emerged for three of the four forms studied. Data from the fourth form did not confirm this hypothesis but indicated a possible interaction between the technical science factor and passage location. After more forms are studied as a follow-up to this hypothesis, experimental research in which several confounding factors can be controlled is needed.

Sentence Completion Items. The main factor that appears to be related to DIF on Sentence Completion items is the distinction between nonscience or surface science references and true science references. Items containing true science references tend to be more difficult for females. This finding was replicated across the four forms studied.

Findings from this research emphasize the perplexing nature of factors thought to contribute to DIF. This situation points to a need for experimental research in which passage features (such as technical versus nontechnical science) and item features (such as true versus

surface science) can be controlled in variable sections of the SAT. Two important variables to control are passage location and item type (particularly inference items pertaining to technical science passages). As an aid to understanding DIF on the experimental items, it would also be interesting to control for other examinee background variables, such as course work in science.

Before experimental research is conducted, it will also be useful to verify the findings from this study on additional test forms, particularly test forms containing technical science reading passages and true science sentence completion items. If a follow-up study is done, attention should be paid to the dimensionality of the matching variable used in DIF analyses. Our study used the total score on the SAT-verbal sections as the matching criterion. However, SAT-verbal dimensionality analyses (Dorans and Lawrence 1987) suggest that it may be more reasonable to match examinees on the Reading subscore instead of on the total SAT-verbal score.

REFERENCES

Dorans, N. J., and E. M. Kulick. 1983. *Assessing unexpected differential item performance of female candidates on SAT and TSWE forms administered in December 1977: An application of the standardization approach.* ETS Research Report 83-9. Princeton, N.J.: Educational Testing Service.

Dorans, N. J., and E. M. Kulick. 1986. Demonstrating the utility of the standardization approach to assessing unexpected differential item performance on the Scholastic Aptitude Test. *Journal of Educational Measurement* 23:355–68.

Dorans, N. J., and I. M. Lawrence. 1987. *The internal construct validity of the SAT.* ETS Research Report 87-35. Princeton, N.J.: Educational Testing Service.

Wendler, C. L. W., and S. T. Carlton. 1987. An examination of SAT-verbal items for differential performance by women and men: An exploratory study. Paper presented at annual meeting of American Educational Research Association, Washington, D.C.

APPENDIX A: SELECTED READING COMPREHENSION ITEMS, DISTRACTOR ANALYSES, AND EMPIRICAL CHARACTERISTIC FUNCTIONS*

Table A.1. Reading Comprehension Text for SAT Form 1, Verbal 1, Items 21–25, with Standardized Proportion Difference between Males and Females (shaded)

Each passage below is followed by questions based on its content. Answer all questions following a passage on the basis of what is stated or implied in that passage.

The boiling point of water depends on the pressure of the air above the surface of the liquid. For example, when the barometric pressure is low—in stormy weather or high on a mountain—the boiling point of water is lower than it is in good weather or at sea level.

When a liquid boils, bubbles form beneath its surface. A bubble forming displaces some of the liquid and will rise to the surface. Because the atmosphere is pressing down on the surface, the rising bubble must push upward against the pressure. If the pressure inside the bubble is less than the pressure of the air on the water surface, the vapor will condense and the bubble will be crushed.

There is a definite relationship between the temperature of a liquid and the pressure that the vapor from the liquid (called the vapor pressure) exerts. At room temperature, molecules near the liquid's surface can escape into the vapor above the liquid if they have sufficient kinetic energy. This process is called evaporation. At higher temperatures, molecules have greater kinetic energy and therefore escape more readily from the liquid into the vapor. As more molecules are converted from liquid to vapor, the vapor pressure above the liquid increases.

When the vapor pressure of water equals the atmospheric pressure ("one atmosphere" at sea level), water boils. At one atmosphere, the temperature of boiling water is 100°C. When water vapor inside a bubble in water exerts a pressure of one atmosphere, it can withstand the pressure around it, and it will not be crushed. When the pressure on a water surface exceeds one atmosphere, the water boils at a temperature higher than 100°C. One place where this happens is in a pressure cooker. A pressure cooker's tight-fitting top has a small escape hole in it. A weight on the escape hole must be lifted before heated air or vapor can escape. The weight is usually chosen so that the pressure of the vapor in the cooker must be two atmospheres before the gas can lift this weight and escape. At two atmospheres the boiling point of water is about 121°C. Foods cook almost three times faster at the higher boiling temperature of water in a pressure cooker.

Looking at a different situation, in a total vacuum there is no pressure on the surface of a liquid. There is no external pressure to overcome. Even near 0°C, the vapor pressure of water is still greater than the pressure around the bubble. So, water in a vacuum can boil at a temperature approaching 0°C.

21. The passage generally concerns the

* −.04 (A) effects of air pressure on the boiling behavior of water
.03 (B) distinction between water and vapor at the molecular level
.01 (C) methods by which water's boiling point can be measured
.00 (D) features of boiling used to make cooking more efficient
.01 (E) ability of water to boil at both extremes of barometric pressure

22. It can be inferred from the passage that water's boiling point is lower on the top of a mountain than its boiling point at the mountain's base because the

.00 (A) peak is closer to the sun and receives more heat
* −.12 (B) atmosphere is thinner at higher altitudes
.05 (C) gravitational pull on rising vapor bubbles is lower on the top
.05 (D) kinetic energy of molecules can be raised more rapidly at higher altitudes
.00 (E) sources of fuel can be employed more efficiently on the top

23. It can be inferred from the passage that increasing the weight that covers the escape hole of a common pressure cooker would cause the water inside to

.01 (A) begin vaporization sooner
.10 (B) come to a boil three times faster
.00 (C) burn any food it might surround
* −.11 (D) boil at a higher temperature
.01 (E) fully evaporate earlier

24. All of the following are mentioned in the passage as having an effect on vapor-bubble formation EXCEPT

.00 (A) temperature
−.01 (B) molecular kinetic energy
.02 (C) barometric pressure
.03 (D) distance above sea level
* −.04 (E) volume of water

25. The passage most likely was intended for an audience of

.03 (A) advanced students of hydrodynamics
.00 (B) mechanics in a water purification plant
−.01 (C) food-preparation employees in industry
* −.10 (D) general readers curious about common phenomena
.08 (E) researchers studying technological applications for water

*Throughout Appendix A, asterisks indicate correct answers.

Table A.2. Reading Comprehension Text for SAT Form 3, Verbal 1, Items 21–25, with Standardized Proportion Difference between Males and Females (shaded)

Each passage below is followed by questions based on its content. Answer all questions following a passage on the basis of what is stated or implied in that passage.

Virtually all living cells depend on solar energy, but most can use it only after it has been captured by photosynthesis. In the process of photosynthesis, green plants produce organic compounds and oxygen, which together contain more chemical energy than the carbon dioxide and water from which they were formed. This increase in chemical energy represents converted solar energy. Thus, as the German physician Julius Robert von Mayer put it, "The world of green plants constitutes a reservoir in which the fleeting rays of the Sun are fixed and ingeniously stored for future use." The organic compounds produced by photosynthesis make up more than 90 percent of the dry weight of every plant on Earth, from minute oceanic plankton to the giant redwood trees. Animals—including human beings—obtain the transformed solar energy either by eating plants or by eating other animals that have fed on the plants.

But life's dependence on photosynthesis goes beyond the food chain. The atmosphere of primordial Earth was devoid of oxygen gas and could sustain only simple single-celled organisms whose modest energy needs were satisfied by fermentation, the breakdown of organic compounds in the absence of oxygen. A form of photosynthesis was taking place, but it did not produce oxygen. About three billion years ago, however, it evolved into the form dominant today, in which oxygen is a by-product. This change dramatically altered the course of evolution. When photosynthesis began producing oxygen, the gas gradually mixed into the atmosphere until the air contained enough of it to support cellular respiration, and this change ushered in an era of energy abundance for all living creatures. Cellular respiration released all of the energy stored in foodstuffs, almost twenty times more than fermentation could. The primitive unicellular organisms that lived by fermentation were soon exceeded by the multitude of respiring, multicellular, higher life forms that came to dominate life on Earth.

As oxygen gradually became a permanent part of the atmosphere, it changed the character of life on Earth yet again. Oxygen gas, on absorbing ultraviolet radiation, formed an ozone layer in the stratosphere, shielding the Earth from deadly forms of solar radiation. Only with the ozone shield in place did the surfaces of land and sea become safe in broad daylight for life forms that previously had lived under the ground or in deep water, where they had been able to avoid the Sun's lethal rays.

In addition to our dependence on photosynthesis for food, oxygen gas, and protection from destructive radiation, we use wood and fiber for shelter and warmth, and we power our industrial civilization by coal, oil, and natural gas—the fossilized products of photosynthesis that took place millions of years ago. Photosynthesis thus eminently merits being called the most important biochemical process on Earth.

21. Which of the following titles best expresses the main theme of the passage?

.00 (A) The Uses of Solar Energy
.00 (B) Our Sources of Energy
.00 (C) The Uses of Oxygen
.00 (D) The Food Chain That Supports Life
* .00 (E) The Importance of Photosynthesis

22. It can be inferred from the passage that one condition necessary for the survival of multicellular organisms is

* –.02 (A) the availability of oxygen for cellular respiration
.03 (B) the presence of unicellular organisms
–.02 (C) a continuous source of ultraviolet radiation
.00 (D) access to wood for shelter and warmth
.01 (E) the absence of fermented substances

23. According to the passage, oxygen was introduced into the Earth's atmosphere when

.00 (A) the solar system was first formed
* –.03 (B) a new form of photosynthesis evolved
.00 (C) single-celled organisms first appeared
.01 (D) fermentation replaced respiration
.01 (E) the ozone layer began to absorb ultraviolet radiation

24. All of the following are mentioned in the passage as being direct or indirect consequences of photosynthesis EXCEPT

.05 (A) conversion of solar energy into food
.00 (B) production of atmospheric oxygen
–.01 (C) availability of energy to the organism
.02 (D) protection of life forms from radiation
* –.06 (E) fermentation of organic compounds

25. It can be inferred from information in the passage that a chemical analysis of a cabbage plant would reveal which of the following?

* –.09 (A) Its dry weight consists mostly of organic compounds.
.00 (B) It contains minute amounts of plankton.
.03 (C) It contains less chemical energy than the materials from which it was formed.
.01 (D) Its cells are incapable of processing oxygen.
.01 (E) It contains substantial amounts of ozone.

Table A.3. Reading Comprehension Text for SAT Form 3, Verbal 2, Items 26–28, with Standardized Proportion Difference between Males and Females (shaded)

Each passage below is followed by questions based on its content. Answer all questions following a passage on the basis of what is stated or implied in that passage.

Society has now fairly got the better of individuality; and the danger that threatens human nature is not the excess, but the deficiency, of personal impulses and preferences. In our times, from the highest class of society down to the lowest, everyone lives as under the eye of a hostile and dreaded censorship. Individuals and families do not ask themselves, "What do I prefer, what would suit my character and disposition, what would allow the best and highest in me to have fair play and enable it to grow and thrive?" They ask themselves, "What is suitable to my position, what is usually done by persons of my station and financial circumstances, or [worse still] what is usually done by persons of a station and circumstances superior to mine?" I do not mean that they choose what is customary in preference to what suits their own inclination. It does not occur to them to have any inclination, except for what is customary. Thus the mind itself is bowed to the yoke: even in what people do for pleasure, conformity is the first thing thought of; they exercise choice only among things commonly done: peculiarity of taste, eccentricity of conduct, are shunned equally with crimes: until by dint of not following their own nature, they have no nature to follow: they become incapable of any strong wishes or native pleasures, and are generally without either opinions or feelings of home growth.

Line
(5)
(10)
(15)
(20)
(25)

26. It can be inferred from the passage that the "hostile and dreaded censorship" (lines 6-7) is a specific reference to

* −.01 (A) the overwhelming anxiety about being perceived as different
−.02 (B) the inescapable danger of offending people in power
−.01 (C) the universal sameness of all members of society
−.01 (D) an individual's first introduction to the idea of conformity
.06 (E) a society's fear of criticism by certain individuals

27. From the author's point of view, which of the following should be the most important consideration in making a decision?

−.02 (A) Decisions made earlier in similar situations
−.01 (B) Advice from personal friends
−.01 (C) Widespread trends in dealing with the type of problem
* .08 (D) Opportunities for self-fulfillment
−.02 (E) Chances for social advancement

28. Which of the following features of modern society does the author consider a threat to human nature?

−.02 (A) Government censorship and thought control
* .07 (B) Lack of independence and spontaneity
−.03 (C) Lack of practical interests and goals
.01 (D) Willingness to follow the dictates of a strong leader
−.01 (E) Fear of being accused of illegal behavior

Table A.4. Reading Comprehension Text for SAT Form 4, Verbal 1, Items 26–30, with Standardized Proportion Difference between Males and Females (shaded)

Each passage below is followed by questions based on its content. Answer all questions following a passage on the basis of what is stated or implied in that passage.

One of the most serious violations of civil liberties in the United States was the treatment of Japanese-Americans during the Second World War. In 1942 the army removed all persons of Japanese descent from coastal areas of three West Coast states and moved them inland to "relocation camps" in which conditions were far from satisfactory. The justification given for this drastic step was that it was necessary to prevent spy activities and sabotage in an essential military zone. But two-thirds of those detained were American citizens. In 1944 the Supreme Court finally ruled on the legal issues involved, holding that the mass evacuation had been legal under the war powers of the President, but that American citizens against whom no charge had been filed and whose loyalty was not questioned could not be detained after removal from the military zone. In the meantime, the War Department repealed its ban against the return of loyal citizens to the coast, but certain elements of the public in the West bitterly opposed the return of the Japanese-Americans and no less than fifty-nine acts of violence were committed in attempts to terrorize them. Although the majority of the people on the West Coast condemned such tactics, many of the Japanese-Americans decided to find new homes in parts of the country where they would encounter less prejudice.

In Hawaii between thirty and forty percent of the population was of Japanese ancestry; hence, it was not possible to remove those persons to the mainland, as was actually suggested, without completely disrupting the economic life of the islands. Therefore, Japanese-Americans in Hawaii were not relocated, but the political rights of all citizens, regardless of their race, color, or nationality, were almost completely abolished during the war. Military law replaced civilian government, and army officers discharged the duties normally assigned to judges and police magistrates. Regulations that might have had meaning right after the 1941 Japanese bombing of Pearl Harbor in Hawaii were enforced with military ruthlessness long after the reasons for their adoption had disappeared.

In spite of the violence and wholesale discrimination against Japanese-Americans, thousands of Hawaiian and mainland Americans of Japanese ancestry served with the United States armed services. Japanese-American battle units had outstanding records in the European theater of war to which they were assigned. No Japanese-American was convicted of either sabotage or espionage on the mainland or in Hawaii before or during the war.

26. It can be inferred from the passage that, according to the 1944 Supreme Court decision, it was legal for the United States government to do which of the following?

    I. Remove all people of Japanese ancestry from coastal areas of three West Coast states
    II. Detain in relocation camps people of Japanese ancestry who were not American citizens
    III. Detain in relocation camps American citizens of Japanese ancestry whose loyalty to the United States was questioned

    .04 (A) III only
    −.01 (B) I and II only
    .01 (C) I and III only
    .00 (D) II and III only
    * −.03 (E) I, II, and III

27. The author would be most likely to agree with which of the following statements about the enforcement of strict military regulations in Hawaii?

    −.02 (A) It was justified as long as a state of war existed.
    * −.01 (B) It was reasonable immediately after Pearl Harbor was bombed.
    −.01 (C) It was defensible if the rules were applied to only certain elements of the population.
    .00 (D) It was necessary because Hawaii was economically important.
    .02 (E) It was unavoidable because Hawaii was so politically unstable.

Table A.4.—Continued

28. According to the passage, Japanese-Americans were not removed from Hawaii during the Second World War because

.02 (A) the military authorities on Hawaii were able to detain them in relocation camps
.00 (B) Hawaii was under attack from the Japanese
.00 (C) there was little resentment against them in Hawaii
* .00 (D) they were important to the economic life of Hawaii
−.02 (E) the distance between Hawaii and the relocation camps on the mainland was too great

29. The passage suggests that which of the following statements about American soldiers of Japanese ancestry during the Second World War is true?

.01 (A) They were not involved in combat.
.08 (B) They were under stricter military regulations than other soldiers.
* −.15 (C) They were in segregated battalions.
.01 (D) They could enlist but they were not drafted.
.03 (E) They were kept in relocation camps unless they volunteered for overseas service.

30. Which of the following would be the best title for the passage?

.01 (A) The Role of Japanese-Americans in Hawaii and on the West Coast of the United States
* −.03 (B) Hostility Toward Japanese-Americans During the Second World War
.01 (C) The Second World War: The Japanese-American Response
−.01 (D) American Attitudes During the Second World War
.02 (E) Military Intervention in Hawaii During the Second World War

Table A.5. Reading Comprehension Text for SAT Form 4, Verbal 2, Items 26–28, with Standardized Proportion Difference between Males and Females (shaded)

Each passage below is followed by questions based on its content. Answer all questions following a passage on the basis of what is stated or implied in that passage.

In most of the truest and most fully satisfying fiction written in the United States after the Second World War, the newborn or self-sustaining or orphaned hero is plunged again and again, for his or her own good and for ours, into the spurious, disruptive rituals of the actual world. We may mention especially *Invisible Man*, by Ralph Ellison; J. D. Salinger's *The Catcher in the Rye*; and *The Adventures of Augie March*, by Saul Bellow—novels in which the heroes are willing, with marvelously inadequate equipment, to take on as much of the world as is available to them, without ever fully submitting to any of the world's determining categories. Ellison's hero is a nameless Black, Salinger's an unstable adolescent, Bellow's a rebellious Chicago Jew. But they share in their common aloneness an odd aura of moral priority over the waiting world. Each of them struggles tirelessly, sometimes unwittingly, and often absurdly, to realize the full potentialities of the classic type each represents—the Emersonian "simple genuine self against the whole world." These characters are clearly indebted to the great innocents of nineteenth-century American fiction, but their creators are also able to engender from within their work the hopeful and vulnerable sense of life that makes experience and so makes narrative action possible. To test that sense of life by irony and drama they must create it from within, for they can scarcely find it any longer in the historic world about them.

26. This passage is primarily concerned with

* .01 (A) defining the nature of the hero in certain contemporary fiction
−.01 (B) criticizing the lack of originality in popular twentieth-century fiction
.01 (C) giving examples of the instability of modern society
.00 (D) prescribing the direction that fiction should take
.00 (E) satirizing the inadequate heroes of recent fiction

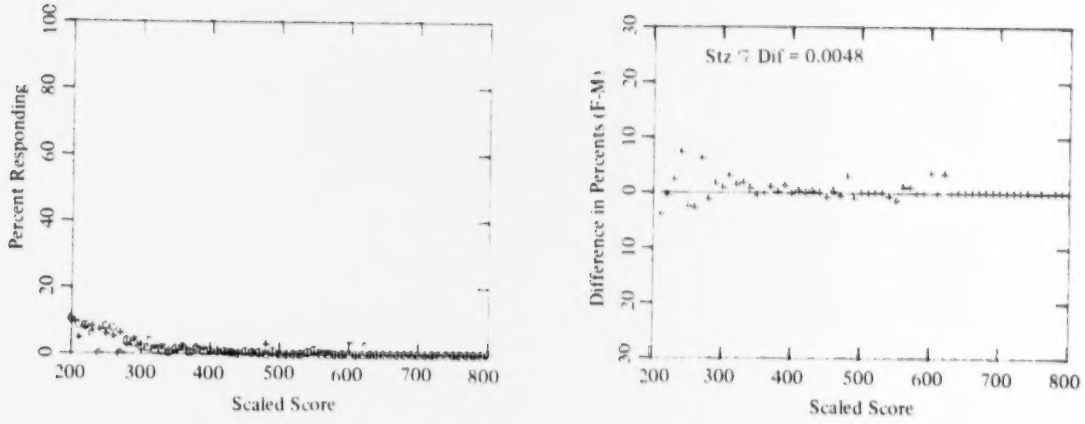
27. The author suggests that all of the characters discussed share the quality of being

−.02 (A) sophisticated
.03 (B) defeatist
* .08 (C) naïve
−.05 (D) impulsive
−.01 (E) intellectual

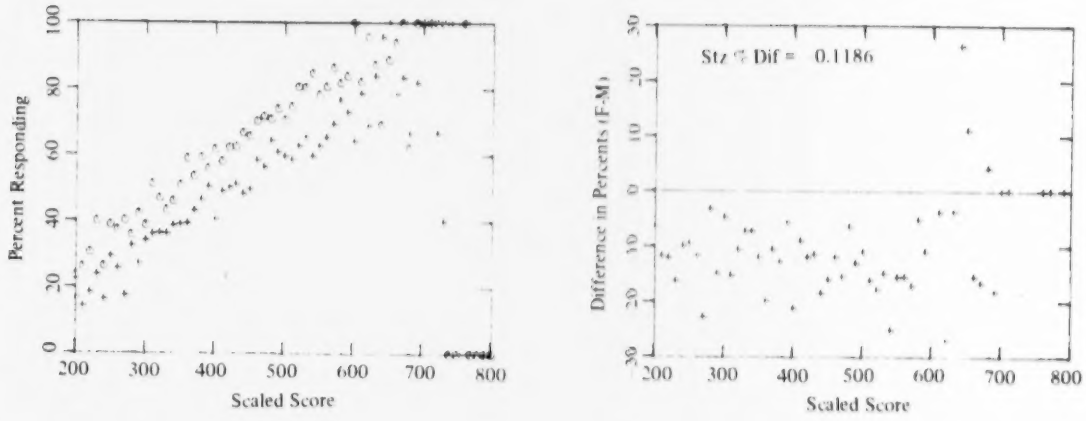
28. The author's attitude toward the fictional characters mentioned seems to be

.01 (A) neutral
.00 (B) contemptuous
−.01 (C) despairing
.00 (D) reverent
* .01 (E) admiring

A



B*



C

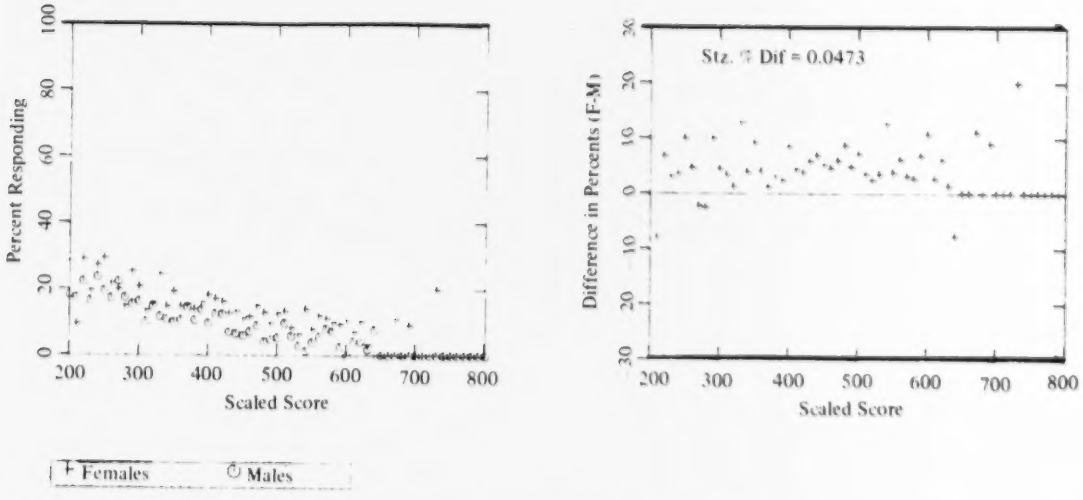
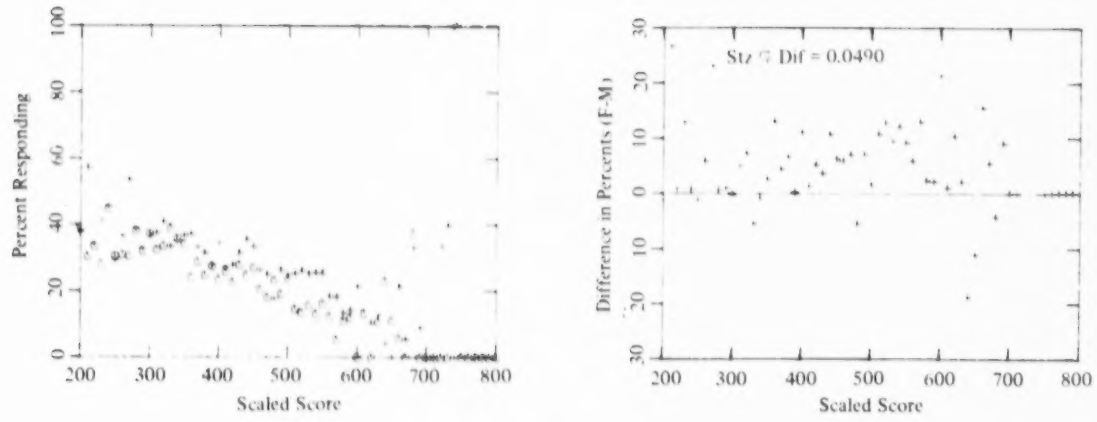
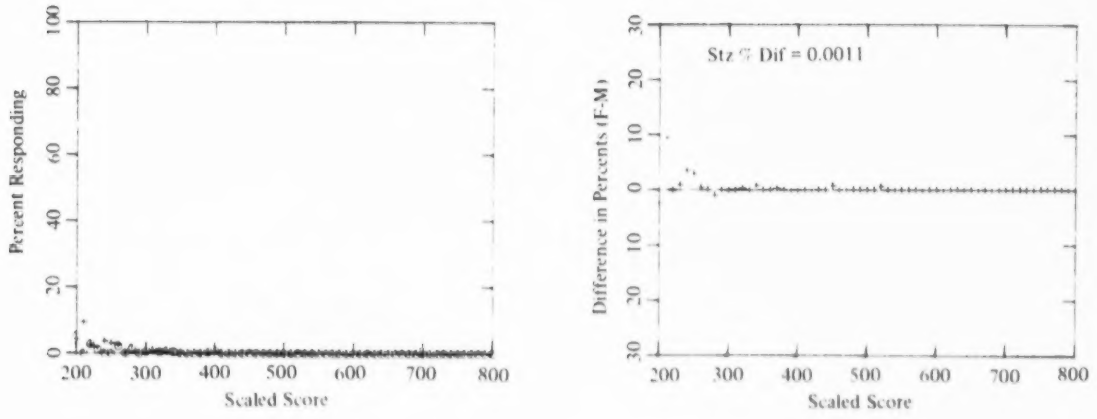


Figure A.1. Empirical Characteristic Functions for Reading Comprehension Item 22 in Form 1, Verbal 1

D



E



OMIT

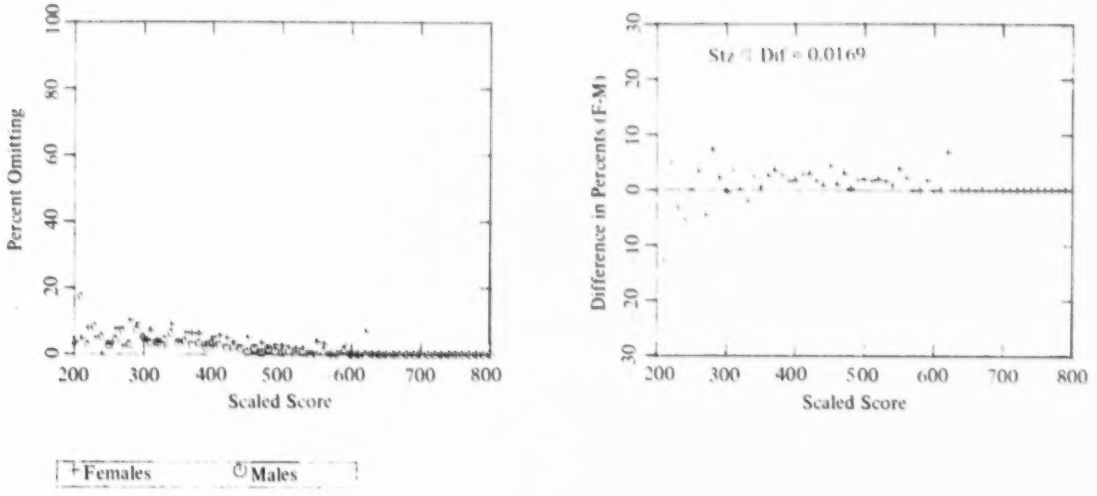
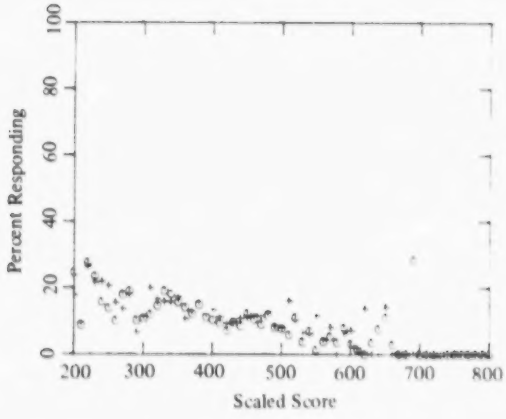
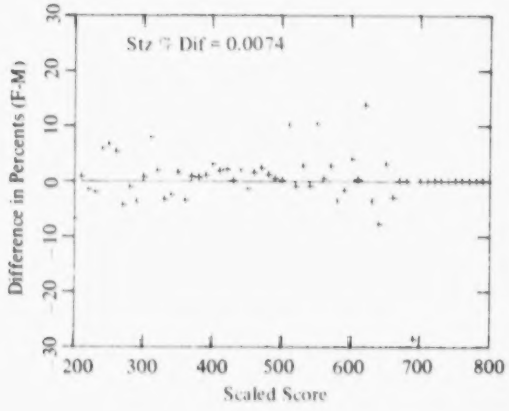


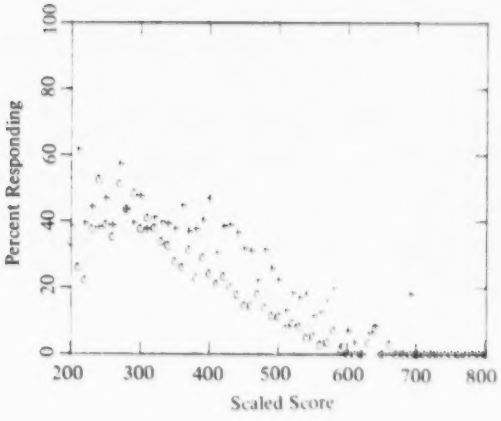
Figure A.1—*continued*

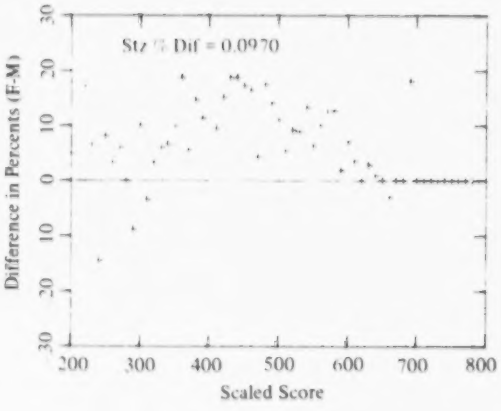
A



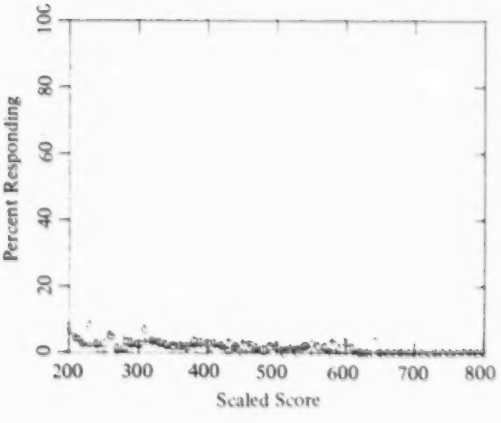


B





C



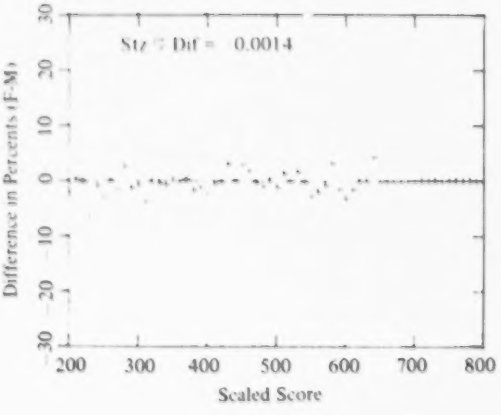
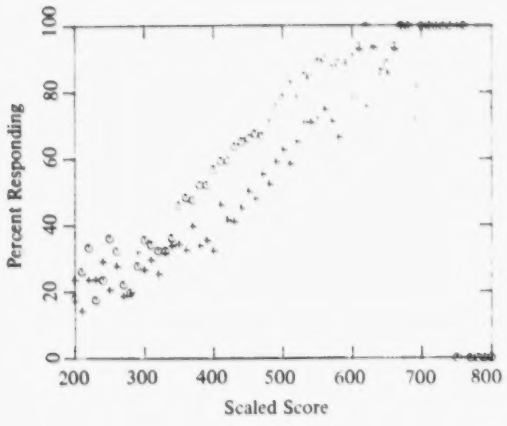
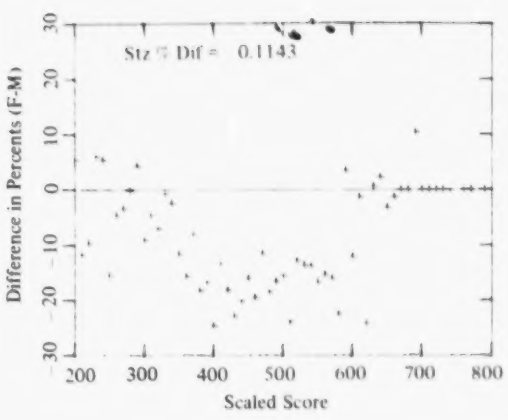


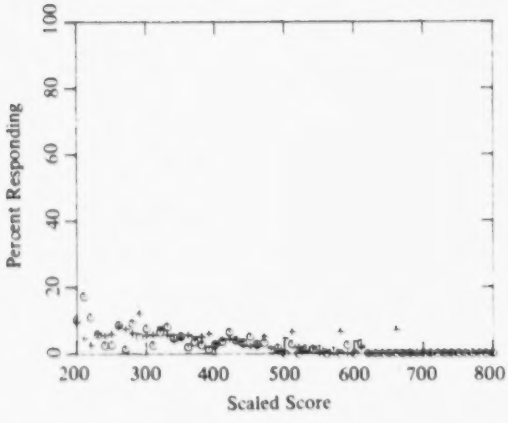
Figure A.2. Empirical Characteristic Functions for Reading Comprehension Item 23 in Form 1, Verbal 1

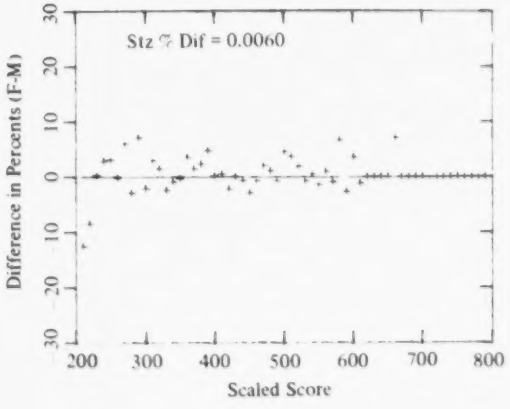
D*



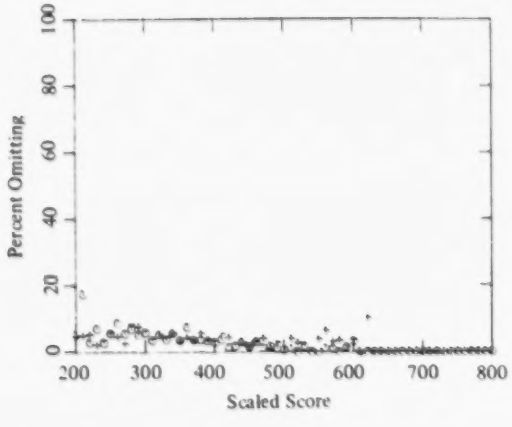


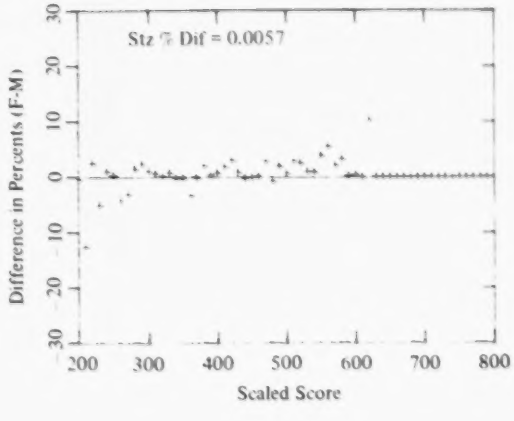
E





OMIT





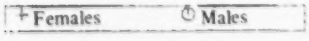
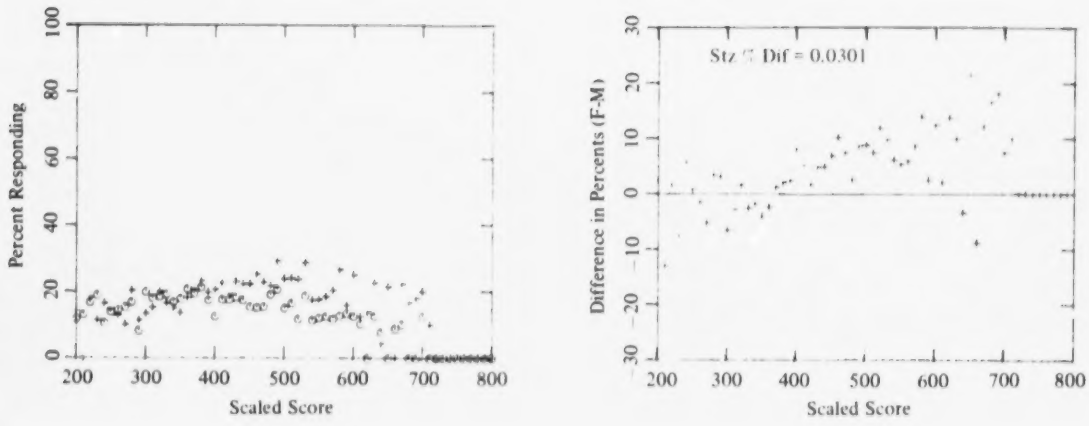
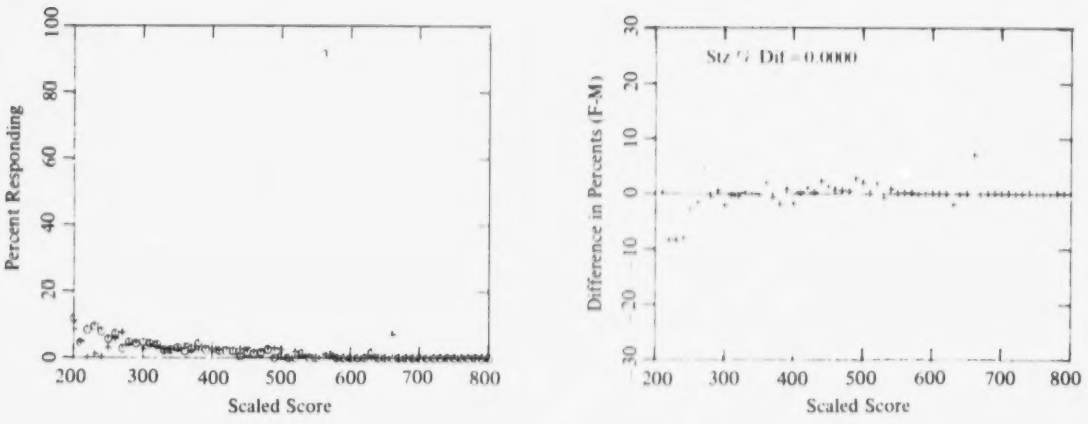


Figure A.2.—*continued*

A



B



C

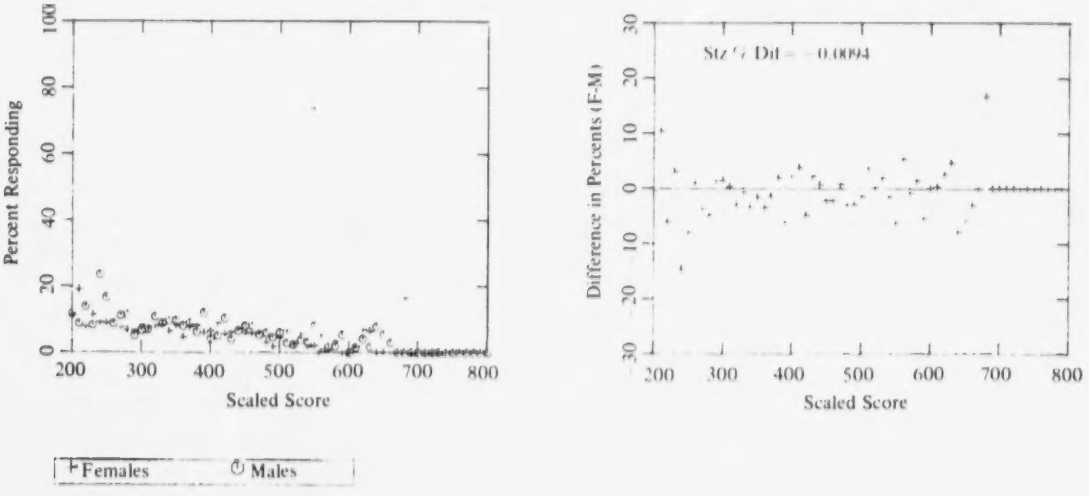
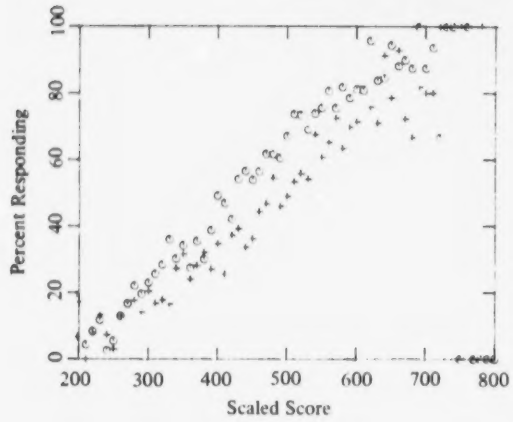
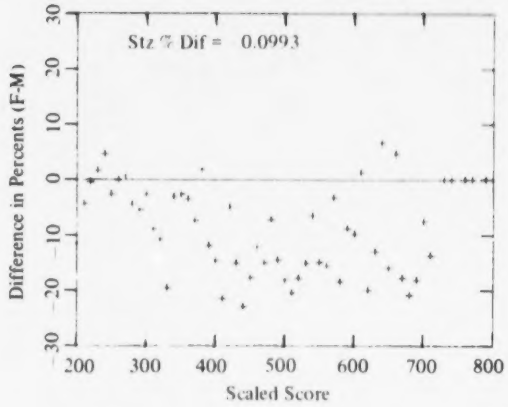


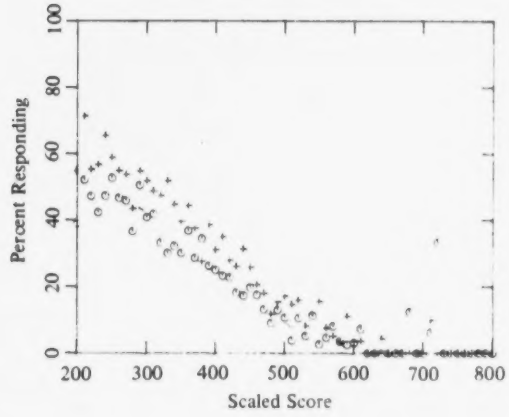
Figure A.3. Empirical Characteristic Functions for Reading Comprehension Item 25 in Form 1, Verbal 1

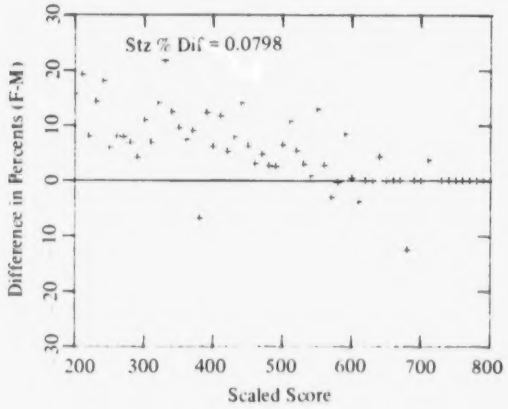
D*



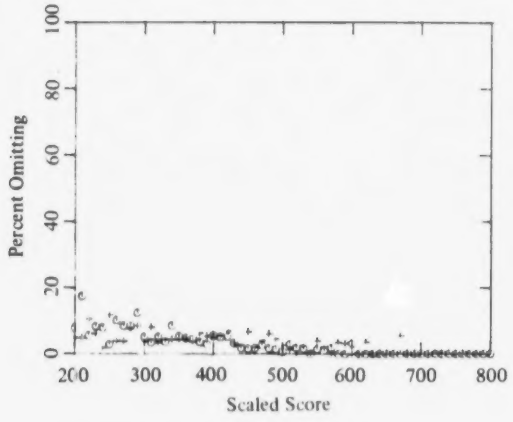


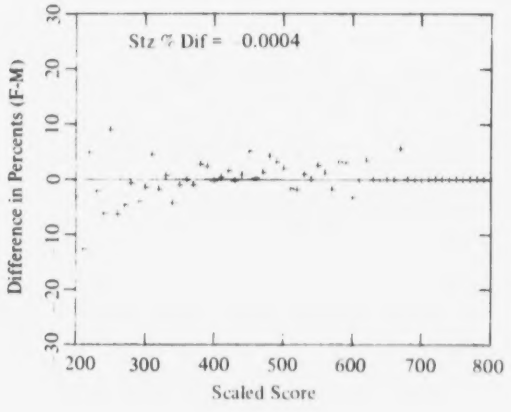
E





OMIT





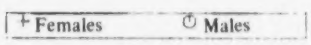


Figure A.3—*continued*

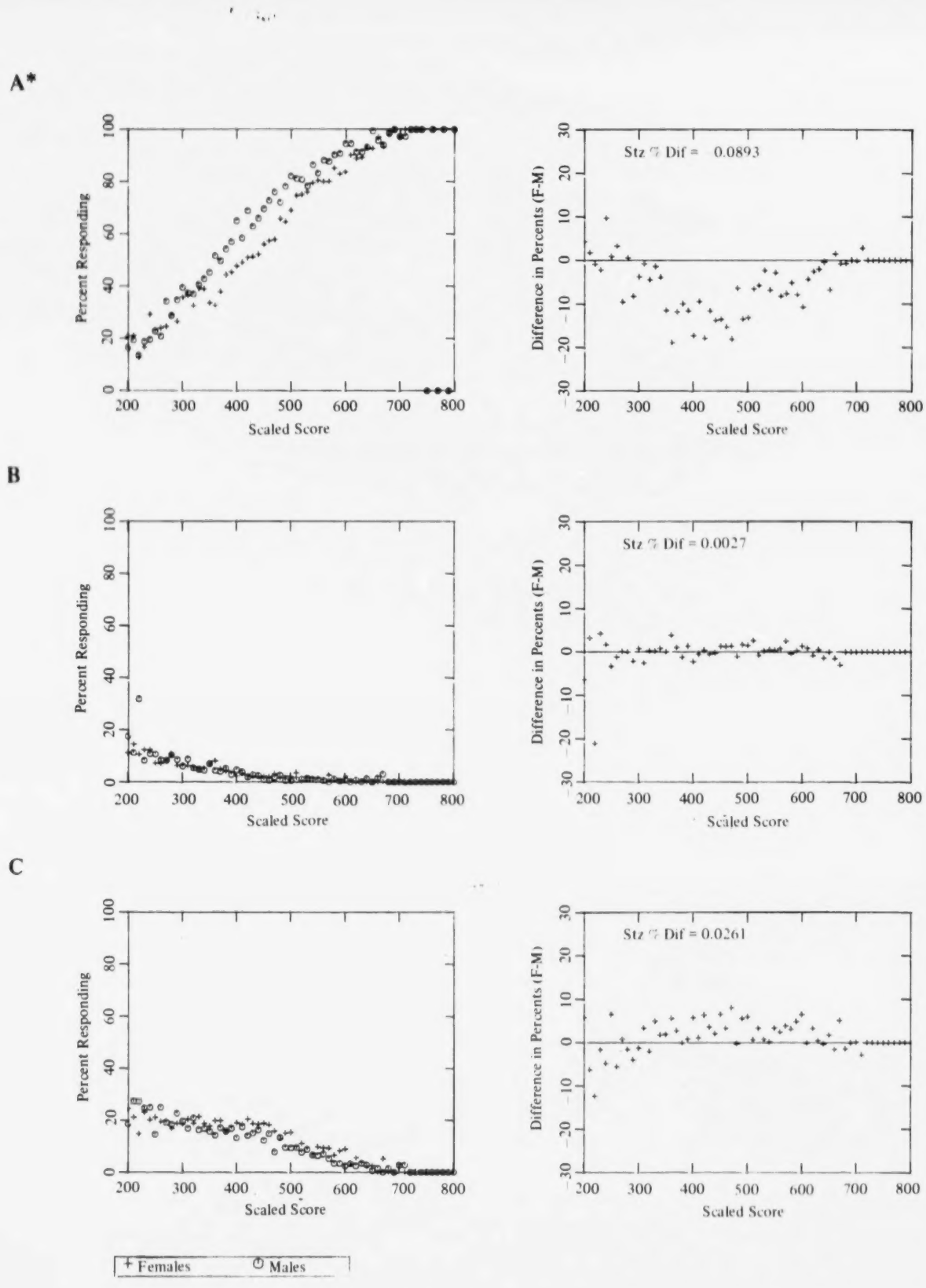
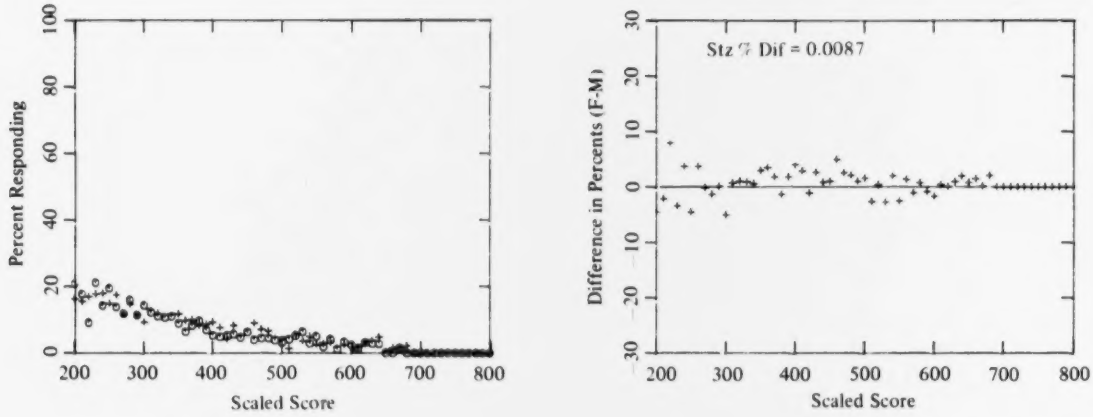
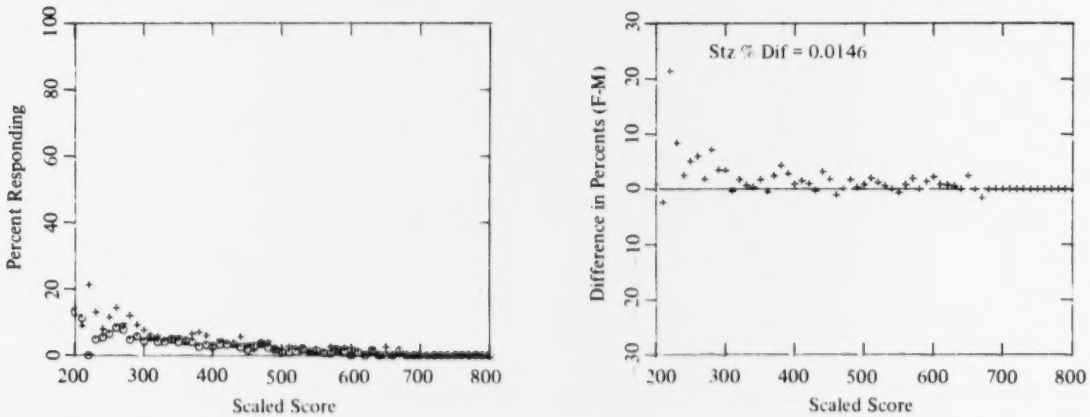


Figure A.4. Empirical Characteristic Functions for Reading Comprehension Item 25 in Form 3, Verbal 1

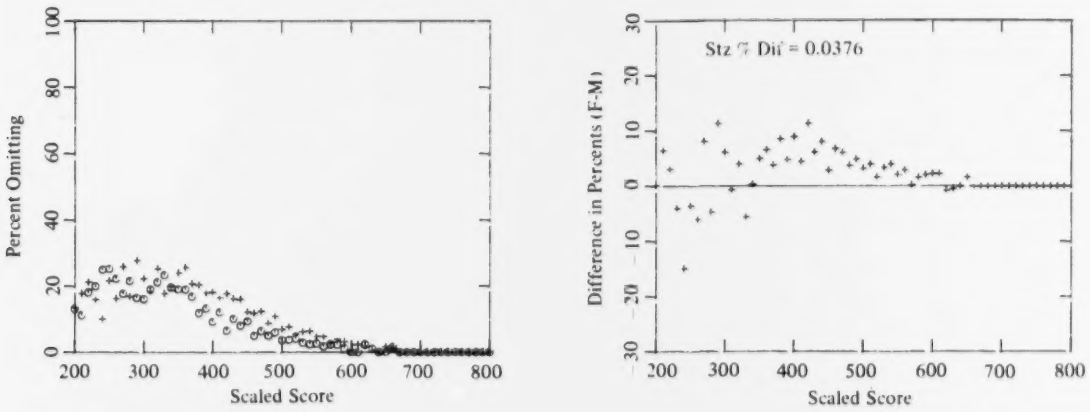
D



E



OMIT



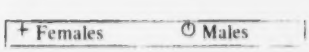
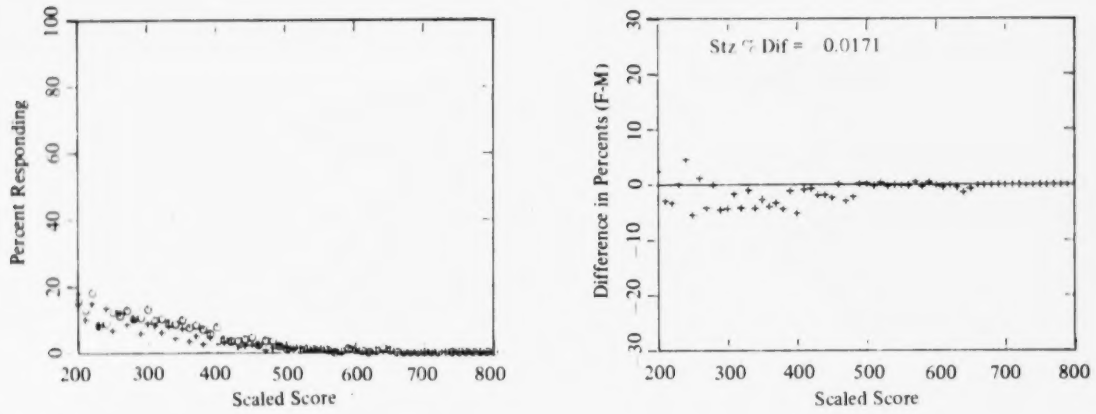
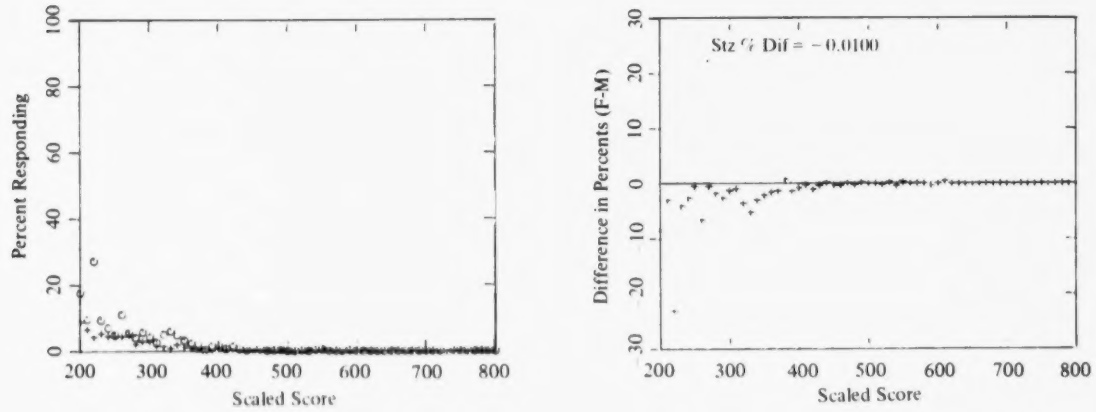


Figure A.4—*continued*

A



B



C

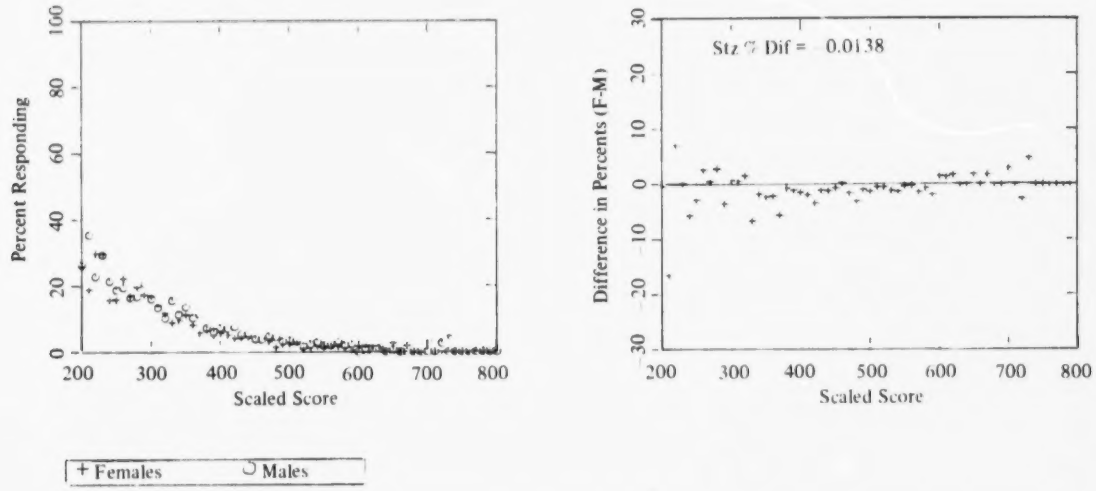
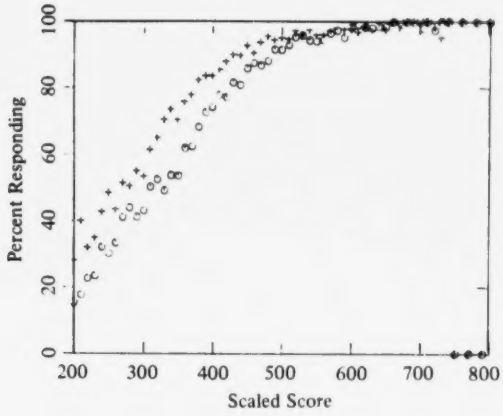
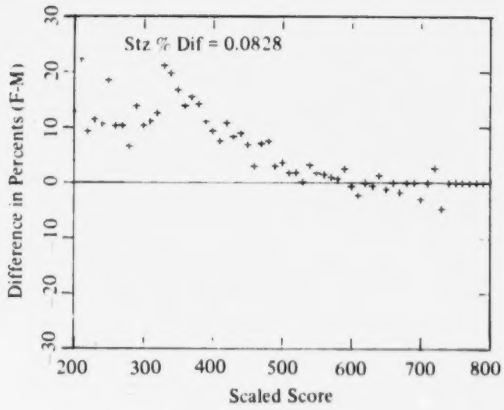


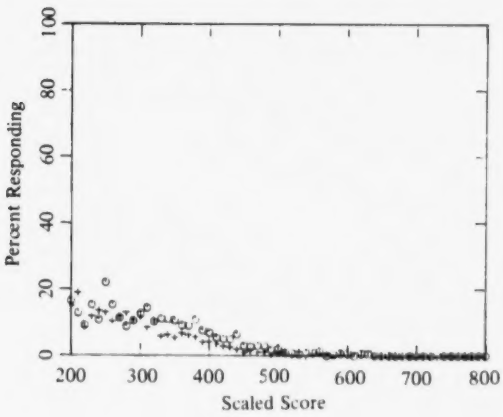
Figure A.5. Empirical Characteristic Functions for Reading Comprehension Item 27 in Form 3, Verbal 2

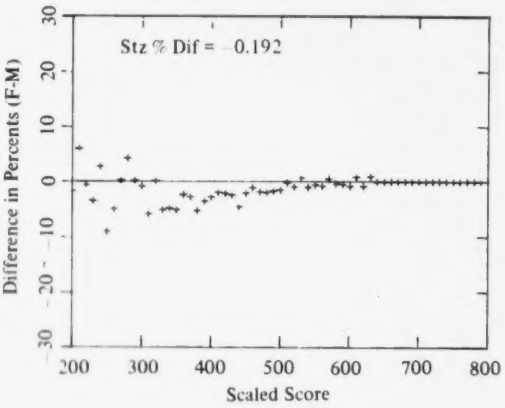
D*



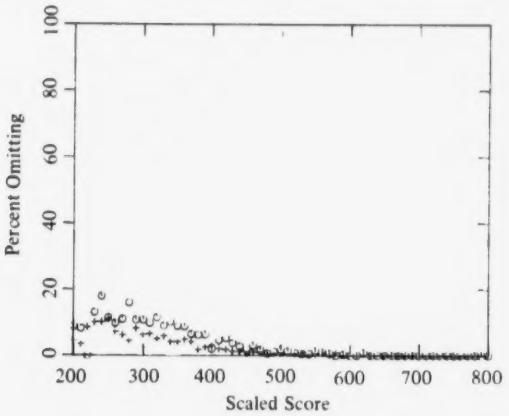


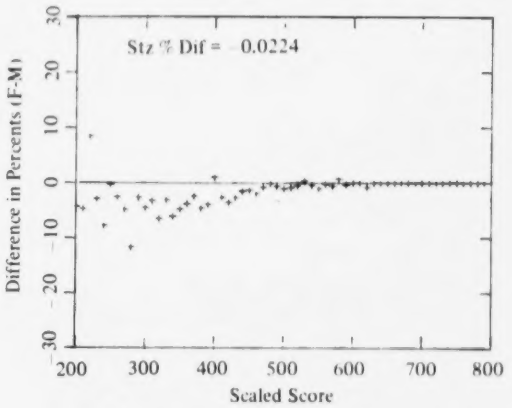
E





OMIT





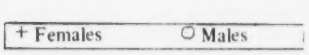
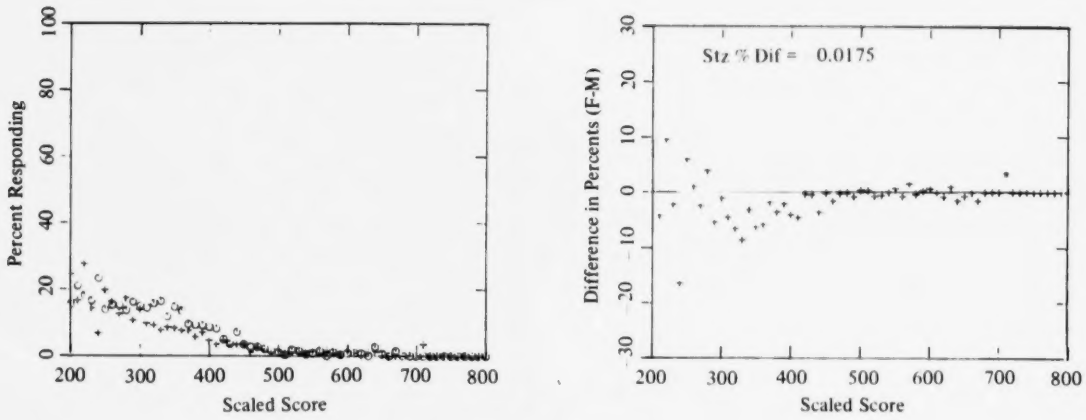
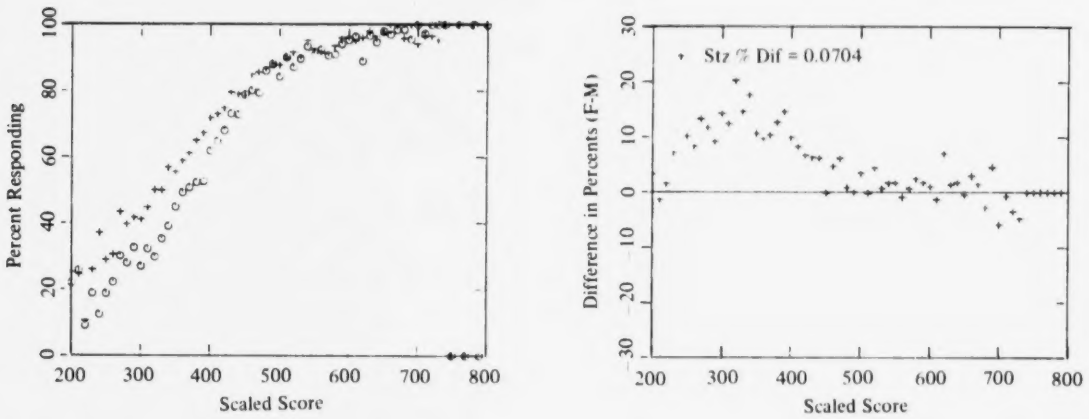


Figure A.5—*continued*

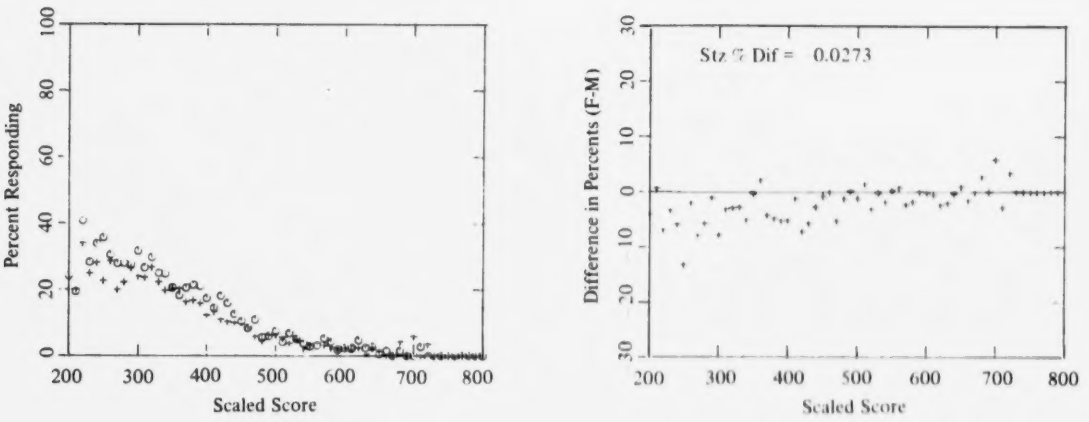
A



B*



C



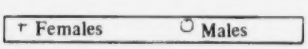
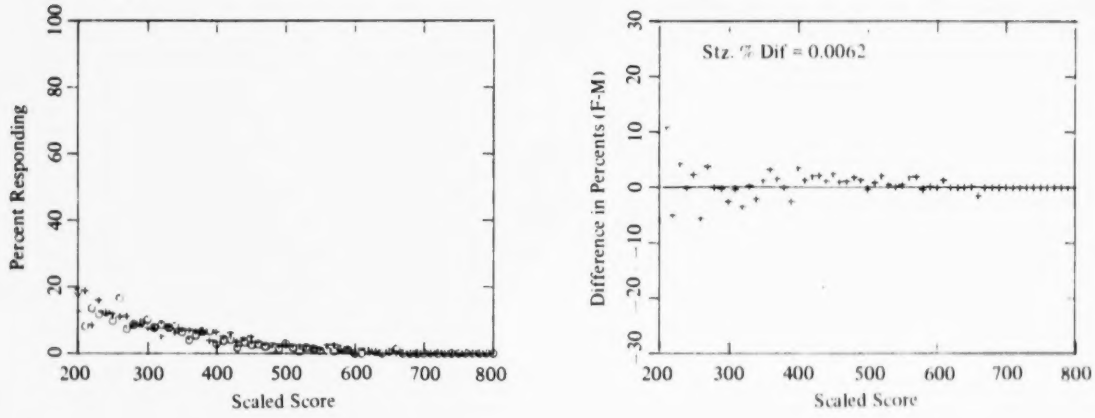
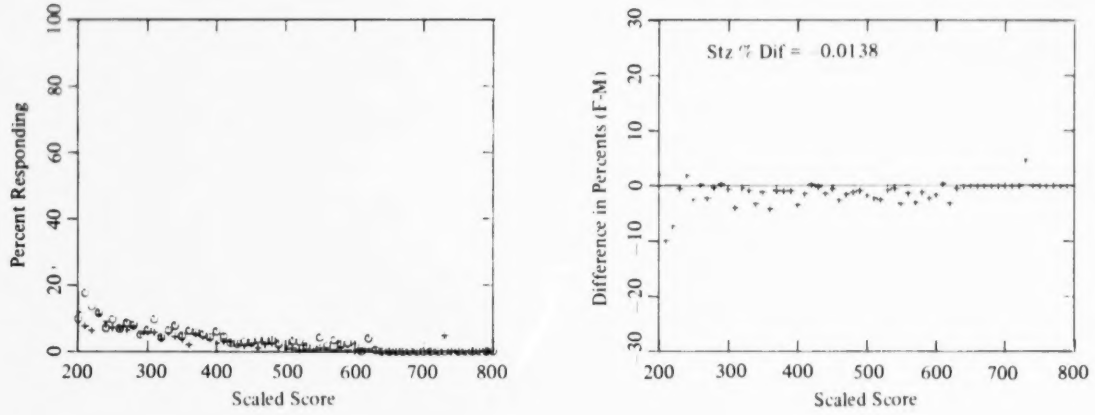


Figure A.6. Empirical Characteristic Functions for Reading Comprehension Item 28 in Form 3, Verbal 2

D



E



OMIT

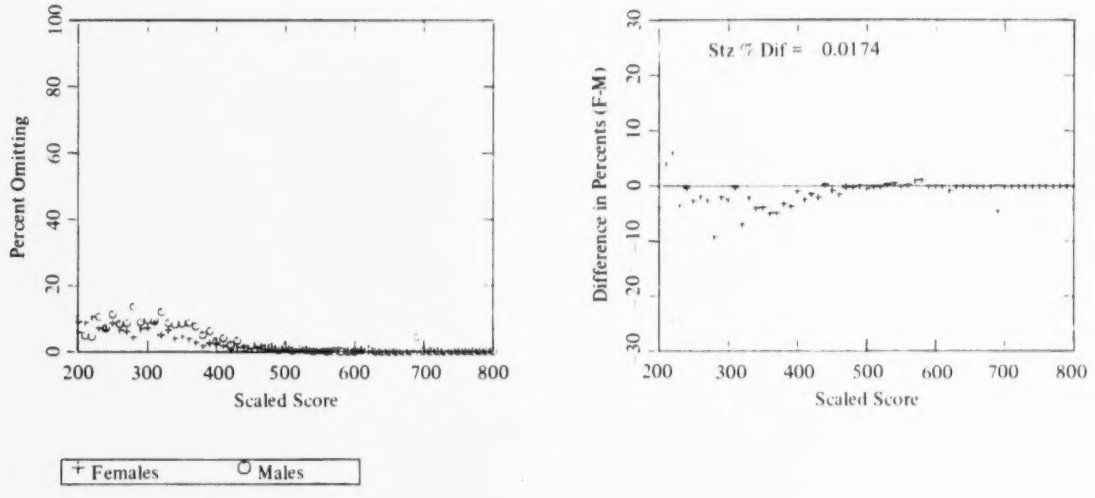
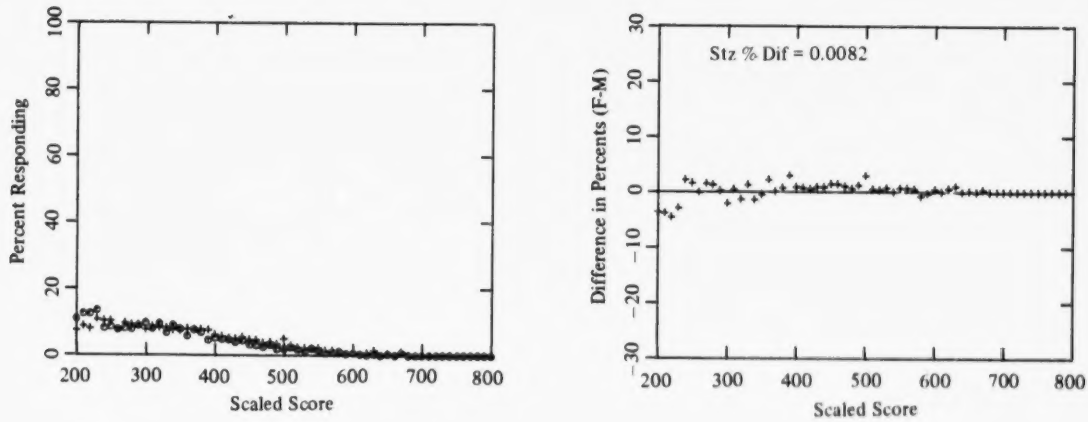
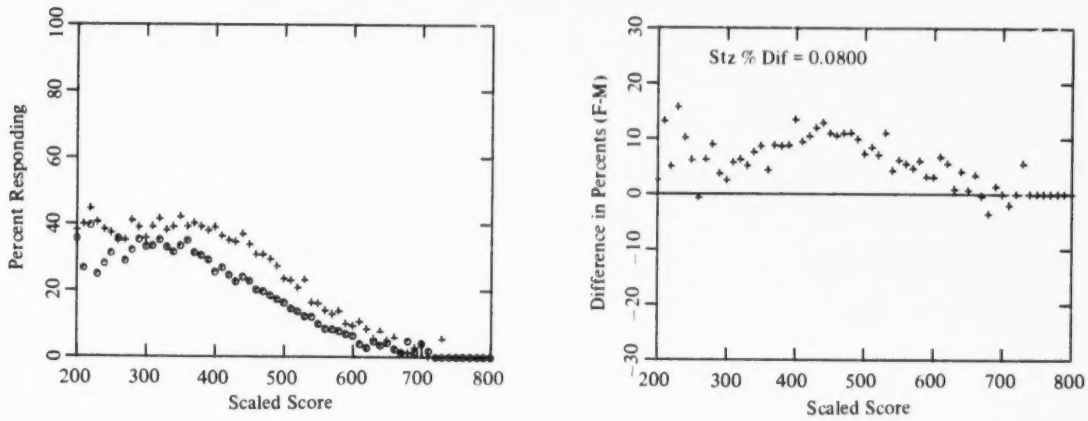


Figure A.6—*continued*

A



B



C*

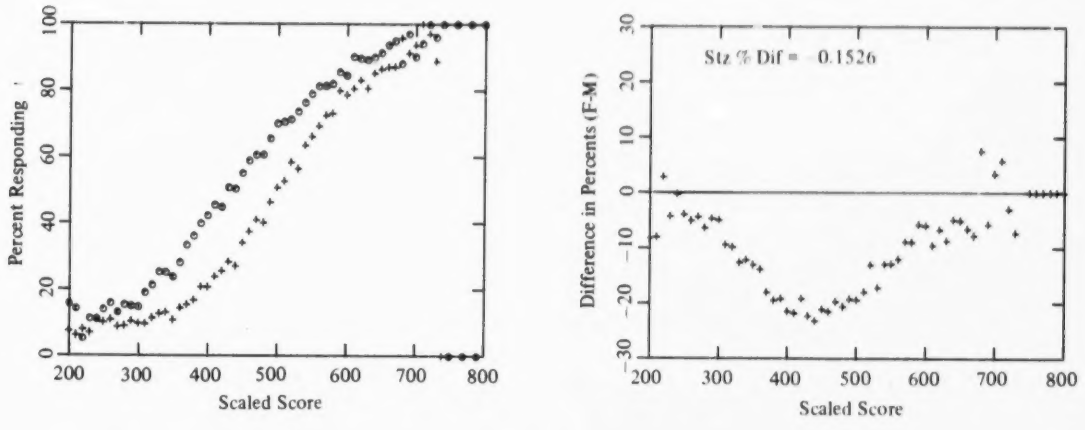
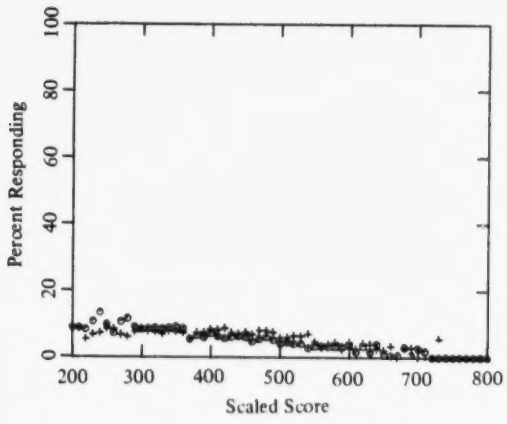
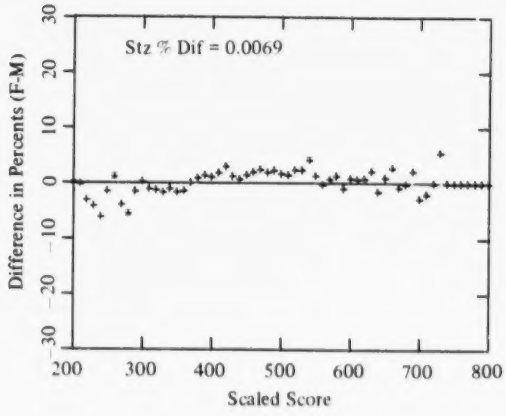


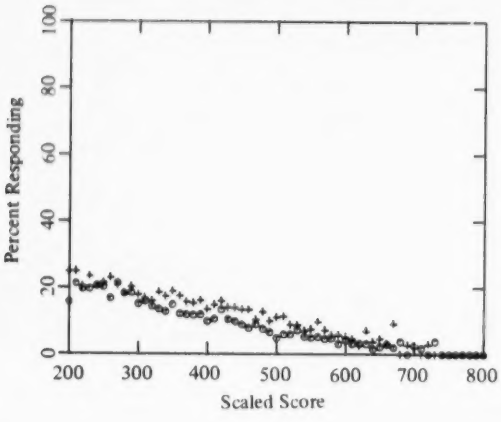
Figure A.7 Empirical Characteristic Functions for Reading Comprehension Item 29 in Form 4, Verbal 1

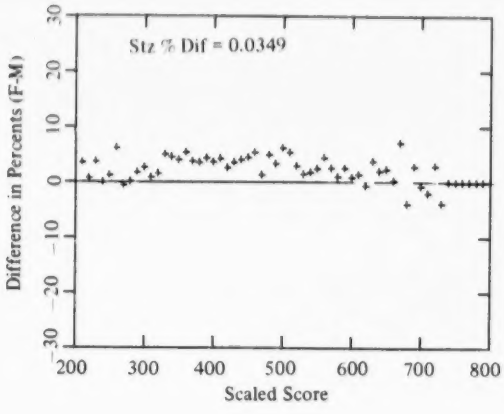
D



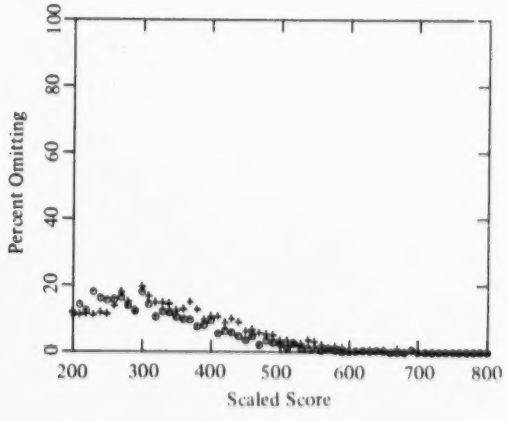


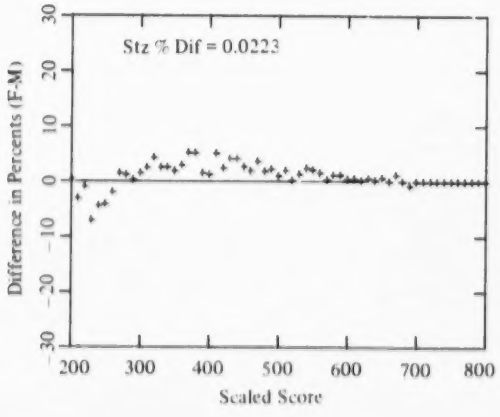
E





OMIT





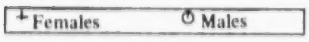
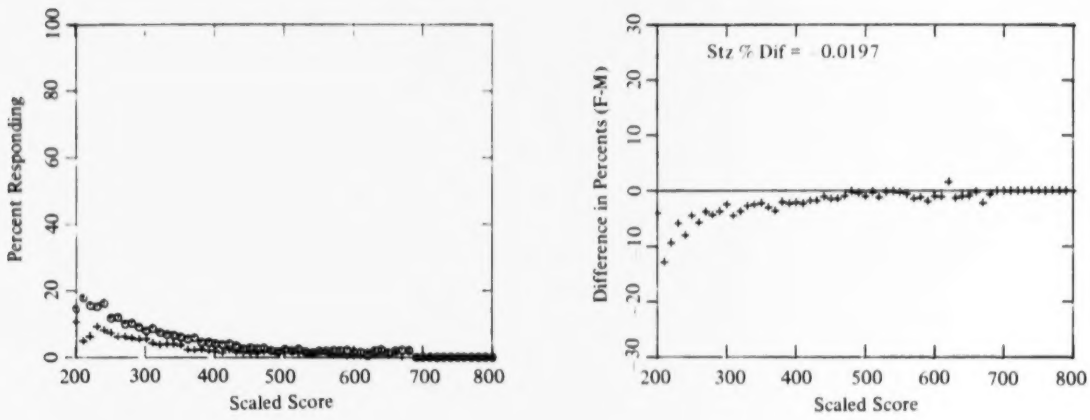
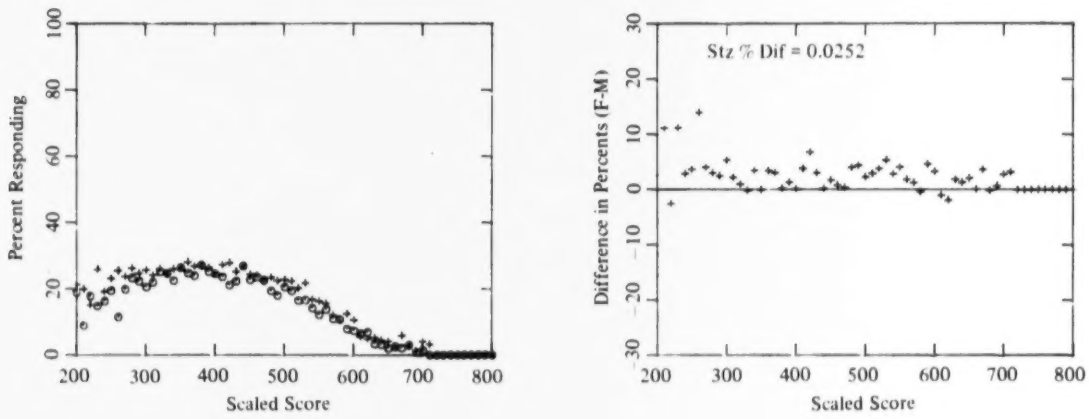


Figure A.7—*continued*

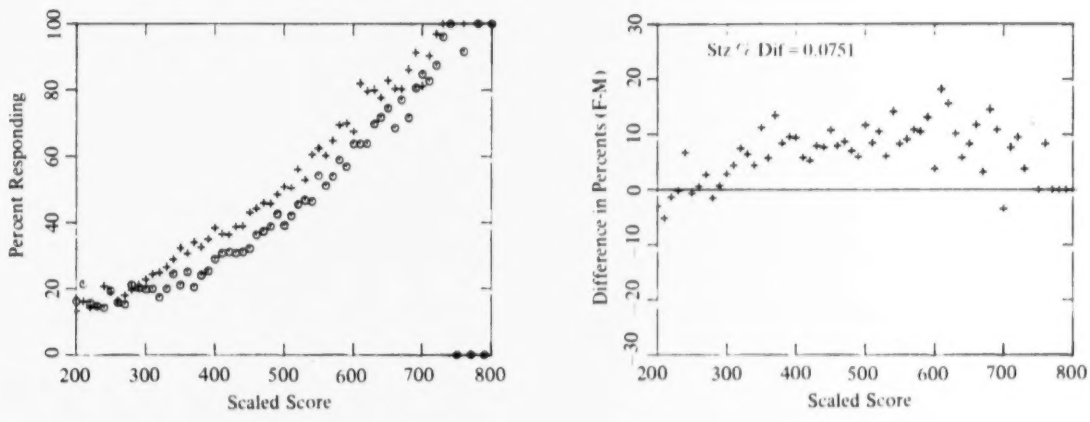
A



B



C*



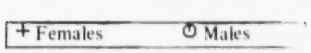
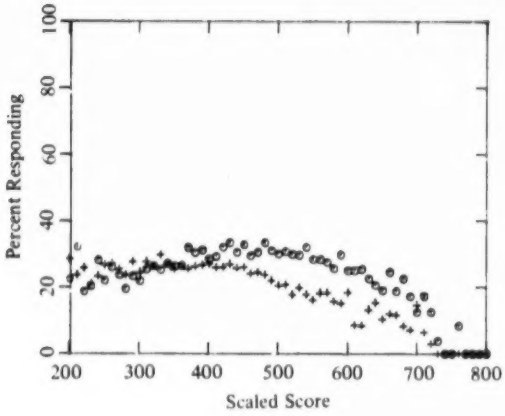
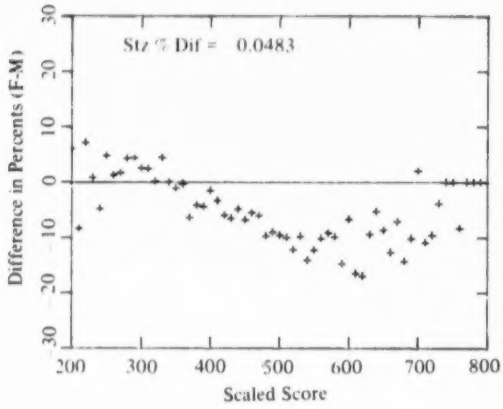


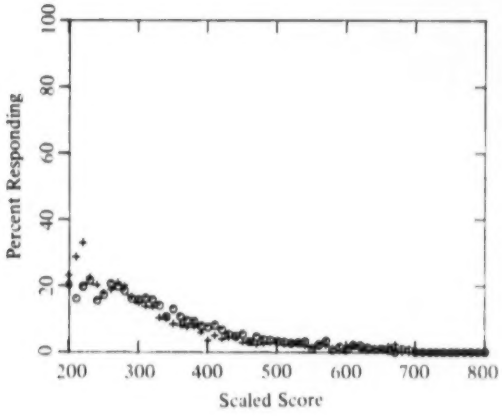
Figure A.8. Empirical Characteristic Functions for Reading Comprehension Item 27 in Form 4, Verbal 2

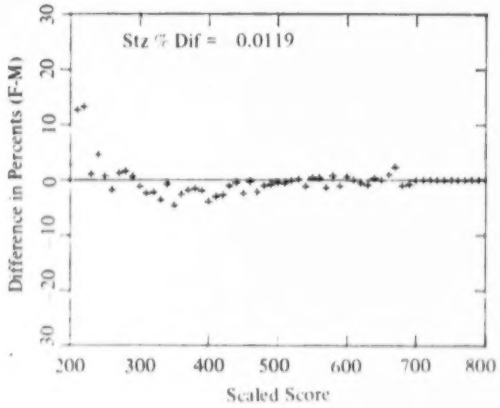
D



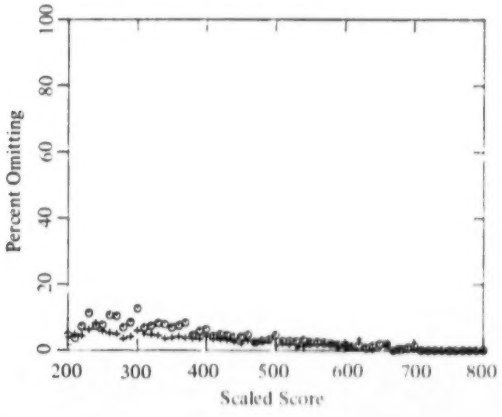


E





OMIT



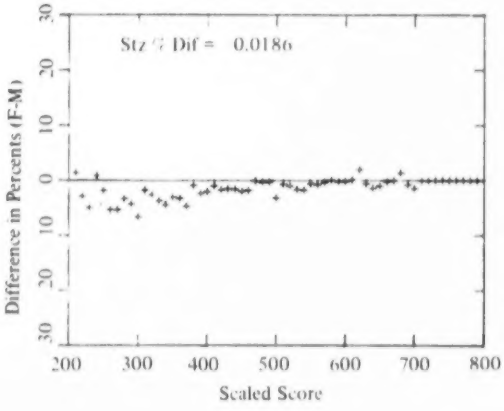


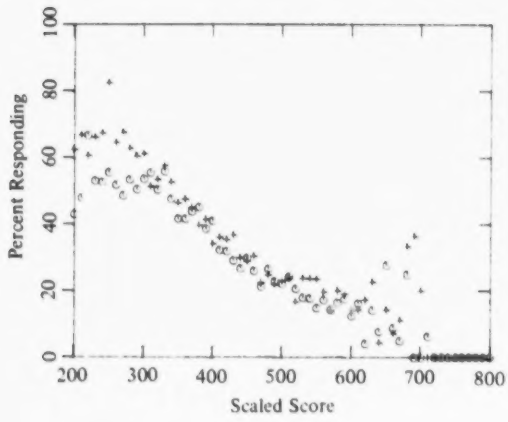
Figure A.8—*continued*

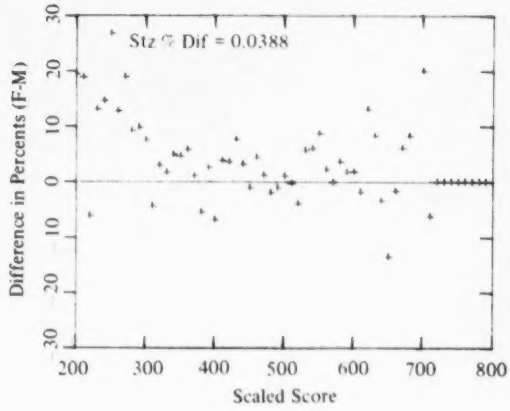
APPENDIX B: SELECTED SENTENCE COMPLETION ITEMS, DISTRACTOR ANALYSES, AND EMPIRICAL CHARACTERISTIC FUNCTIONS

Table B.1. Examples of Sentence Completion Items with High Values of STPDIF

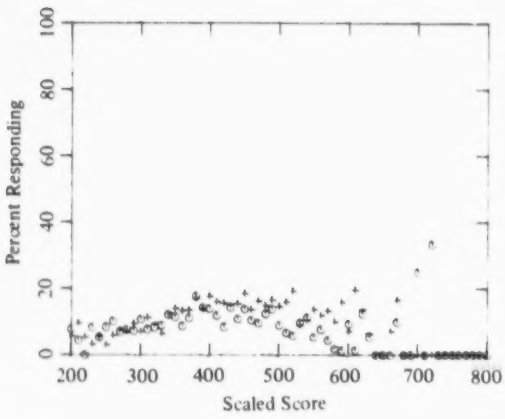
Form	*Item*	*Section*	*Item*		*STPDIF*
1	14	V2	We are always fighting nature's tendency to degrade the ---- and destroy the meaningful, its tendency to surrender life to ---- .		
			A	immortal . . eternity	.04
			B	mutant . . ossification	.02
			C	organized . . decay	−.08 (key)
			D	unchangeable . . stasis	.02
			E	vigorous . . birth	−.01
				OMITTED	.01
3	18	V1	The most ---- means of transportation in the world is the bicycle; indeed, no powered vehicle requires less energy to move as much mass over the same distance.		
			A	grandiose	.02
			B	infallible	.04
			C	efficient	−.08 (key)
			D	engrossing	.01
			E	unstable	.01
				OMITTED	.01
4	20	V1	In the fifteenth century, ancient authority and empirical investigation were ---- ; not until the seventeenth century was a system created by which the two might be reconciled.		
			A	unorthodox	−.08
			B	incompatible	.15 (key)
			C	inadequate	−.04
			D	unintelligible	−.01
			E	interchangeable	−.01
				OMITTED	−.01
4	11	V2	The ---- of trees in East Africa has caused the number of native antelopes to ---- sharply because they can live only where the forest is most dense.		
			A	planting . . decline	.06
			B	destruction . . grow	.00
			C	felling . . decrease	−.06 (key)
			D	abundance . . dip	−.00
			E	protection . . drop	.00
				OMITTED	.00

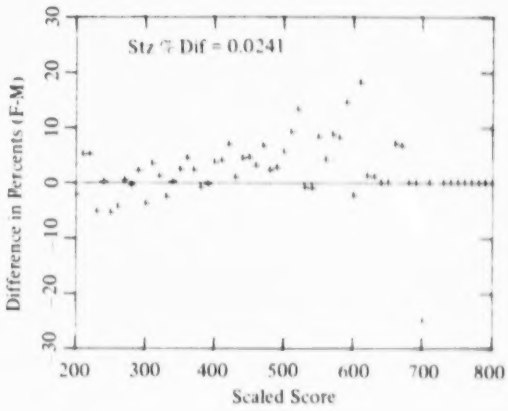
A



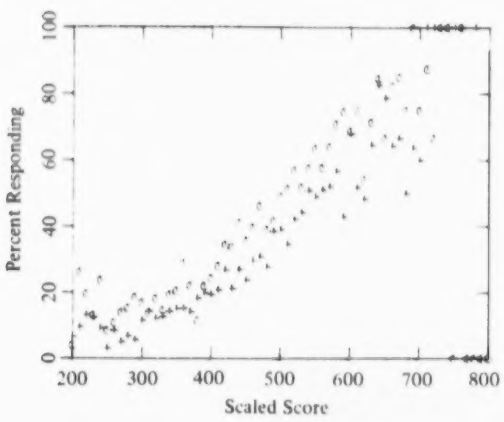


B





C*



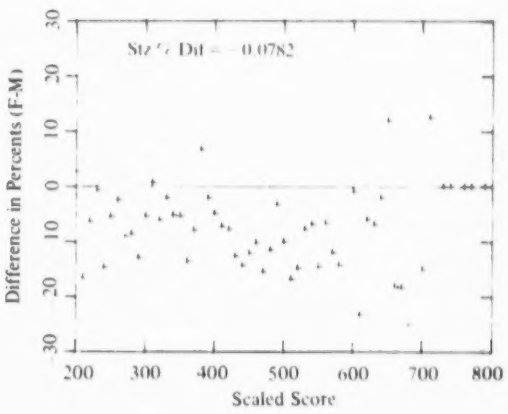
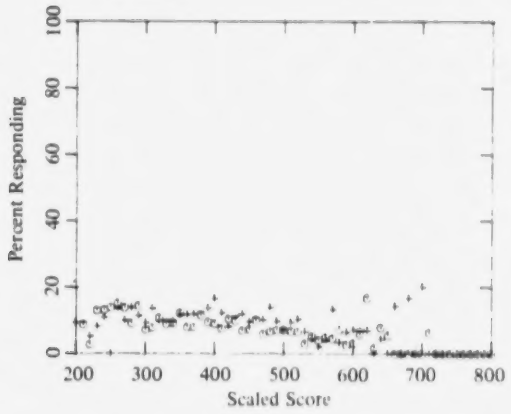
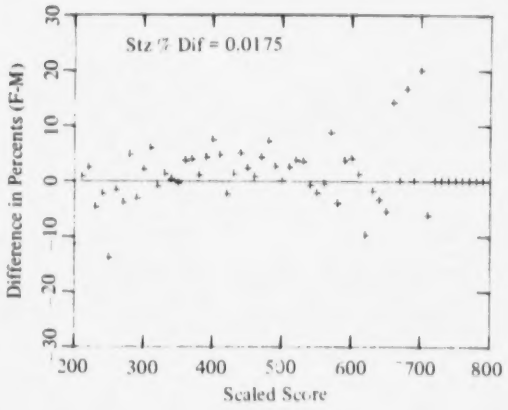


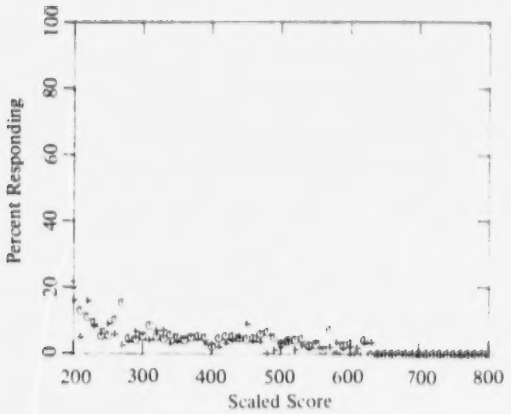
Figure B.1. Empirical Characteristic Functions for Sentence Completion Item 14 in Form 1, Verbal 2

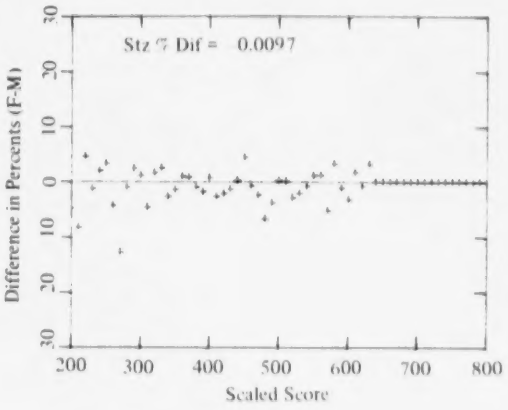
D



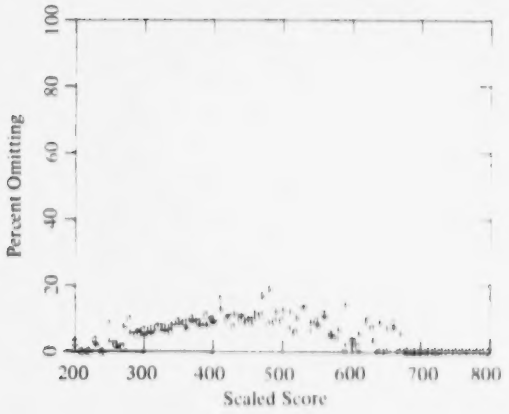


E





OMIT



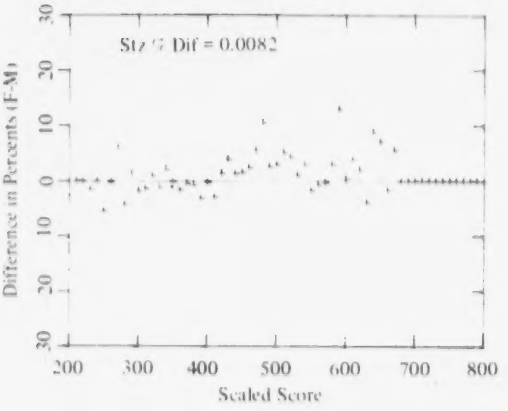


Figure B.1—*continued*

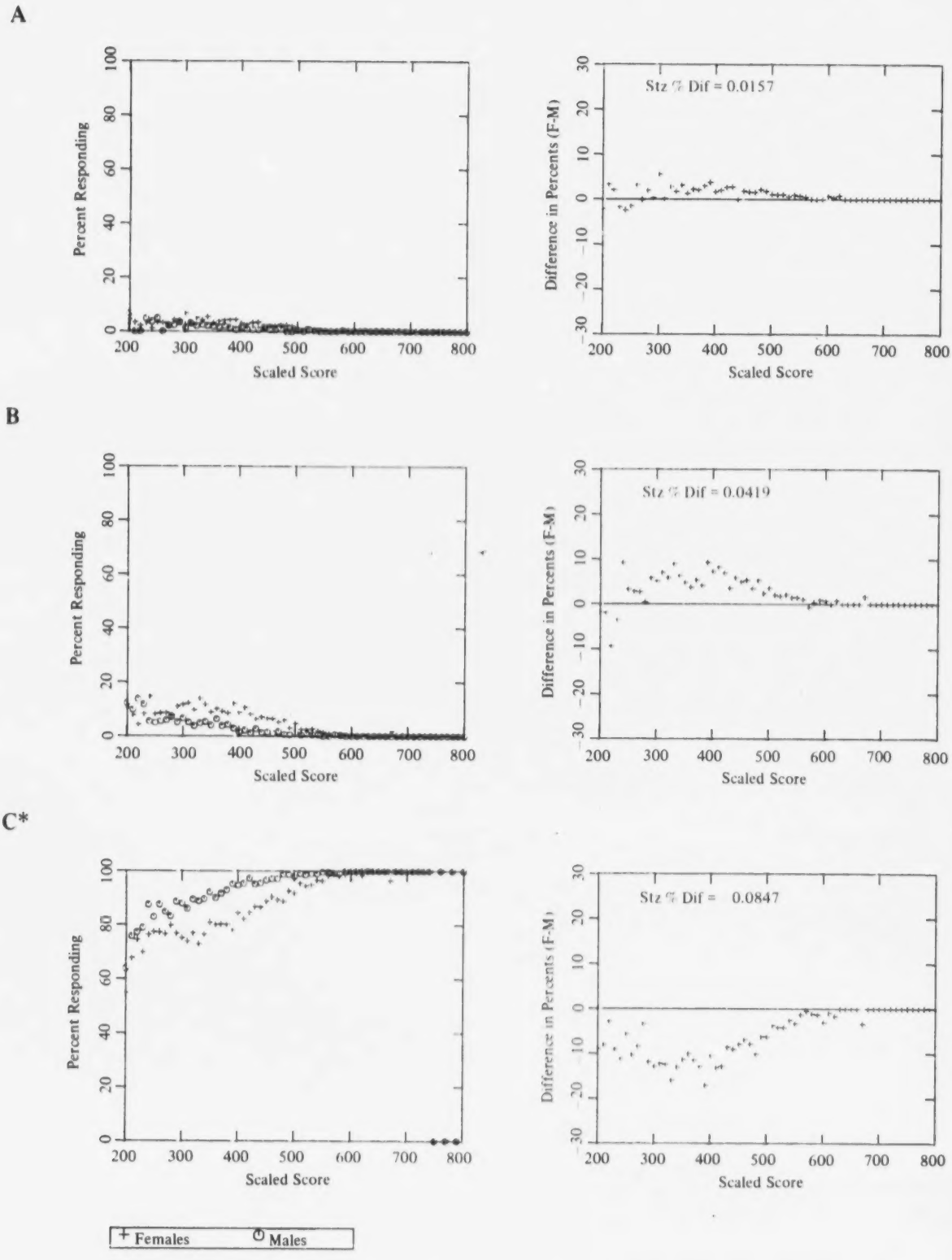
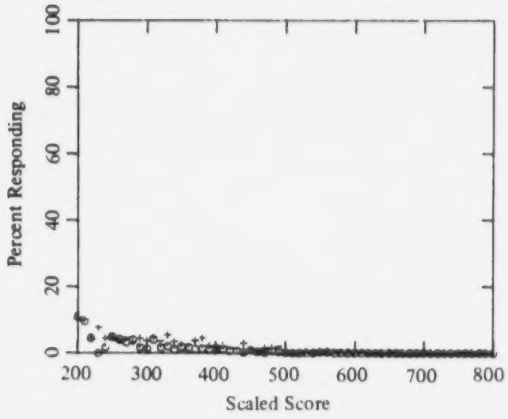
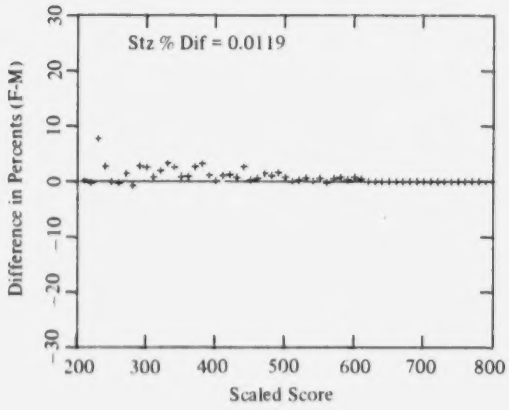


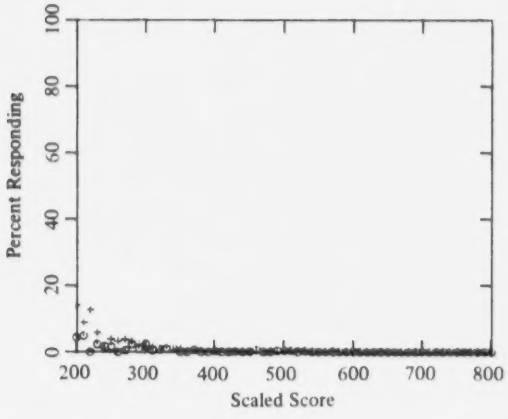
Figure B.2. Empirical Characteristic Functions for Sentence Completion Item 18 in Form 3, Verbal 1

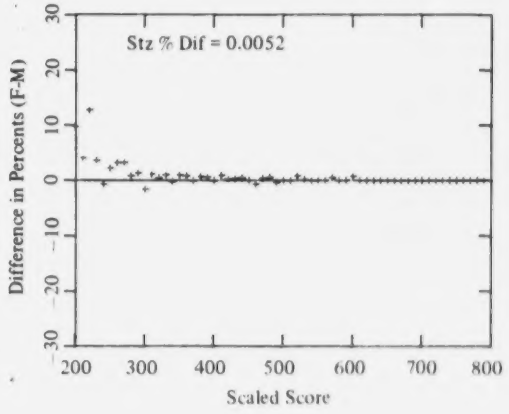
D



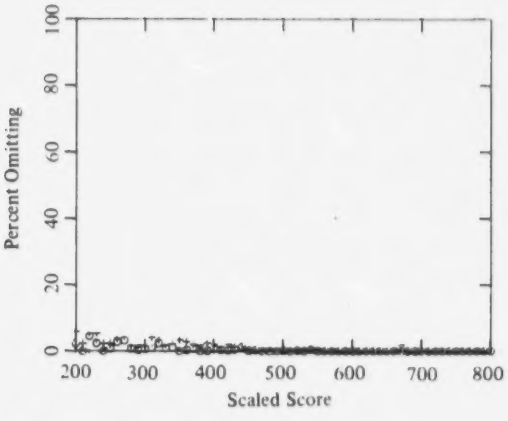


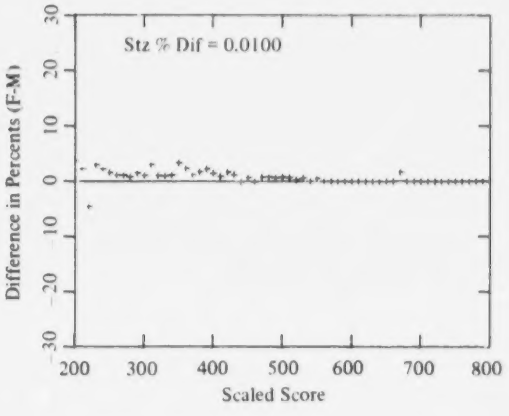
E





OMIT





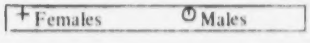
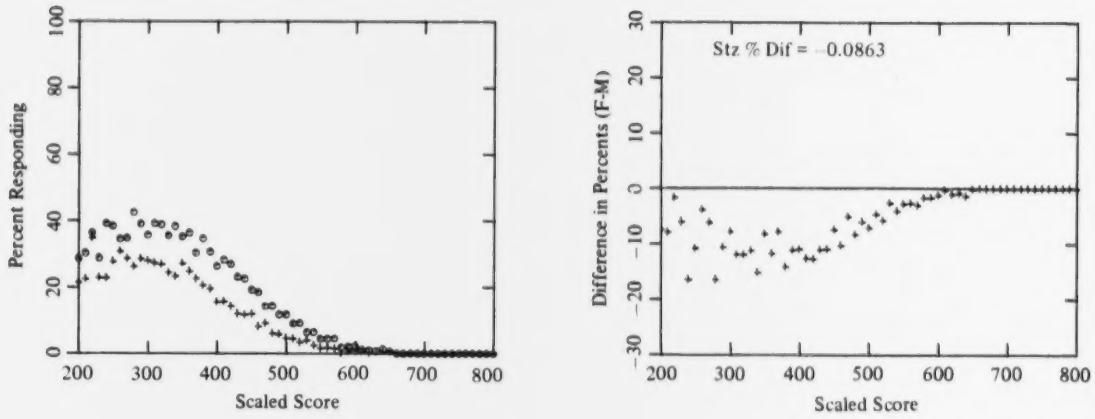
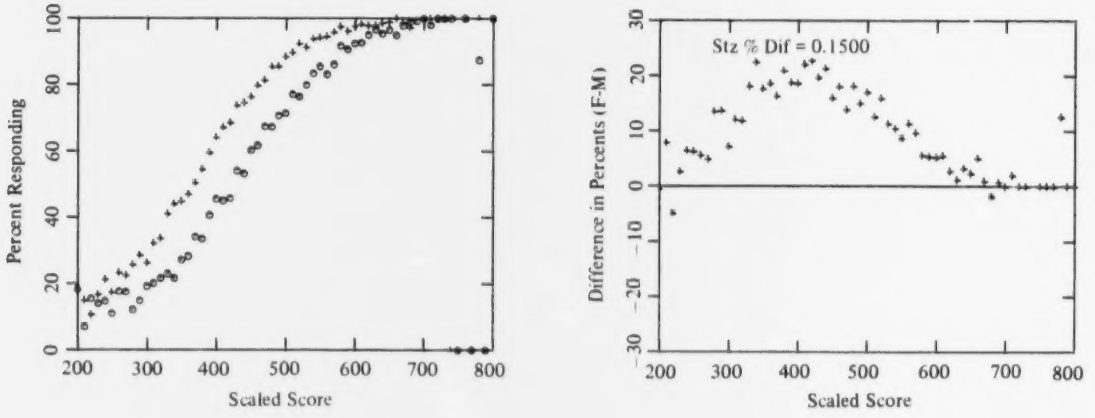


Figure B.2—*continued*

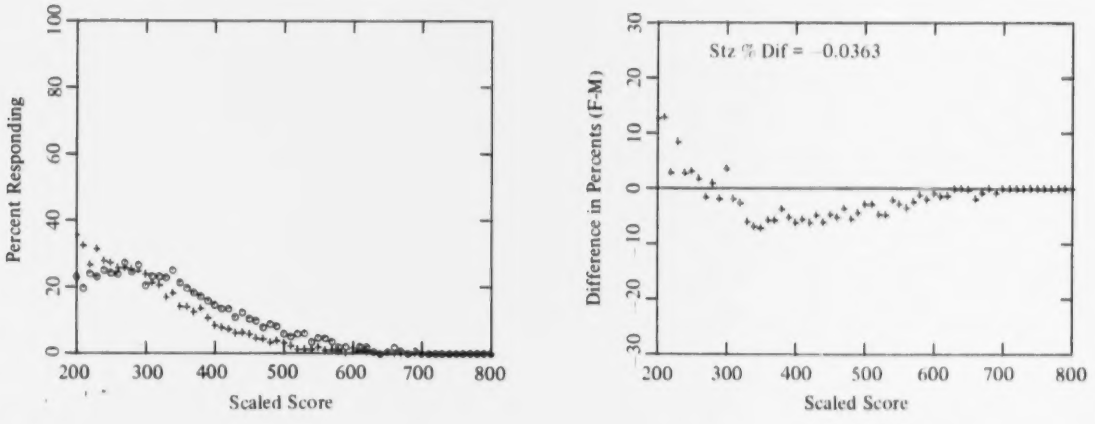
A



B*



C



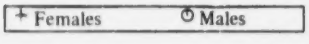
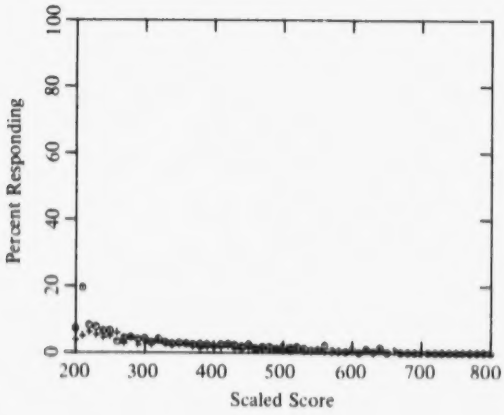
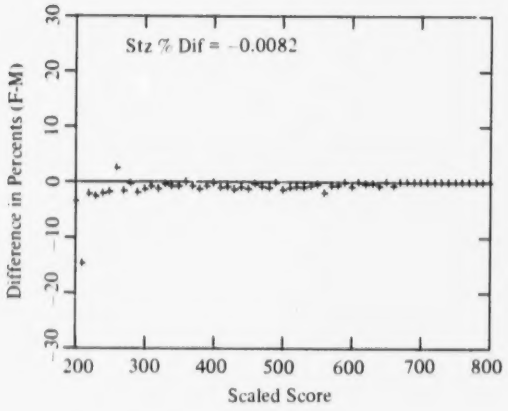


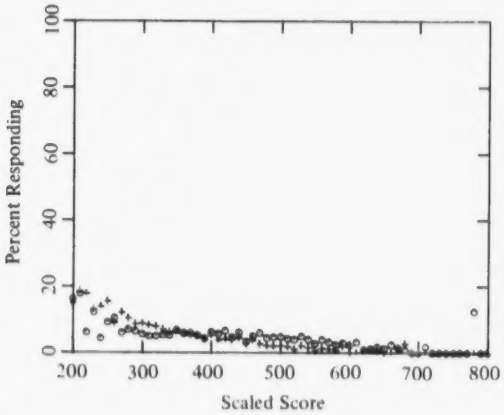
Figure B.3. Empirical Characteristic Functions for Sentence Completion Item 20 in Form 4, Verbal 1

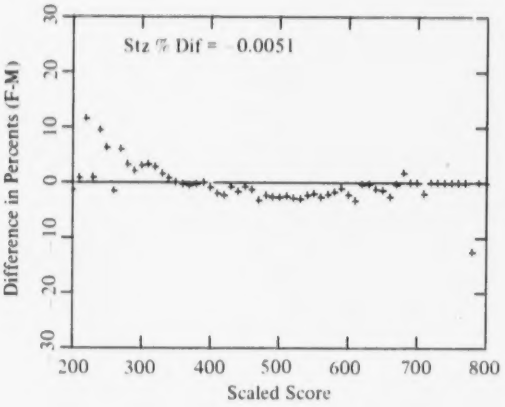
D



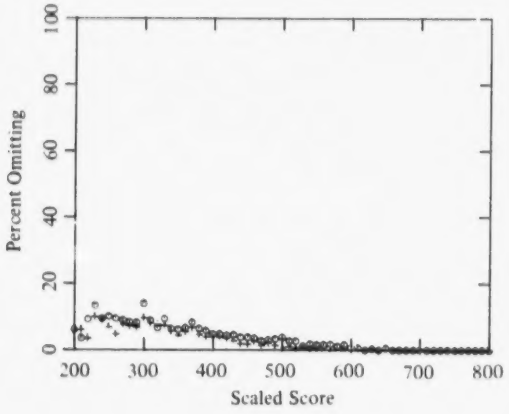


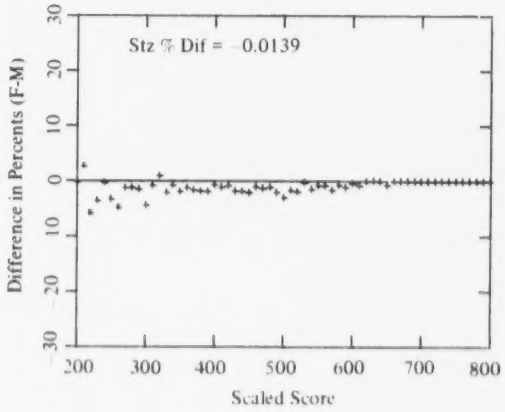
E





OMIT





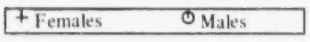


Figure B.3—*continued*

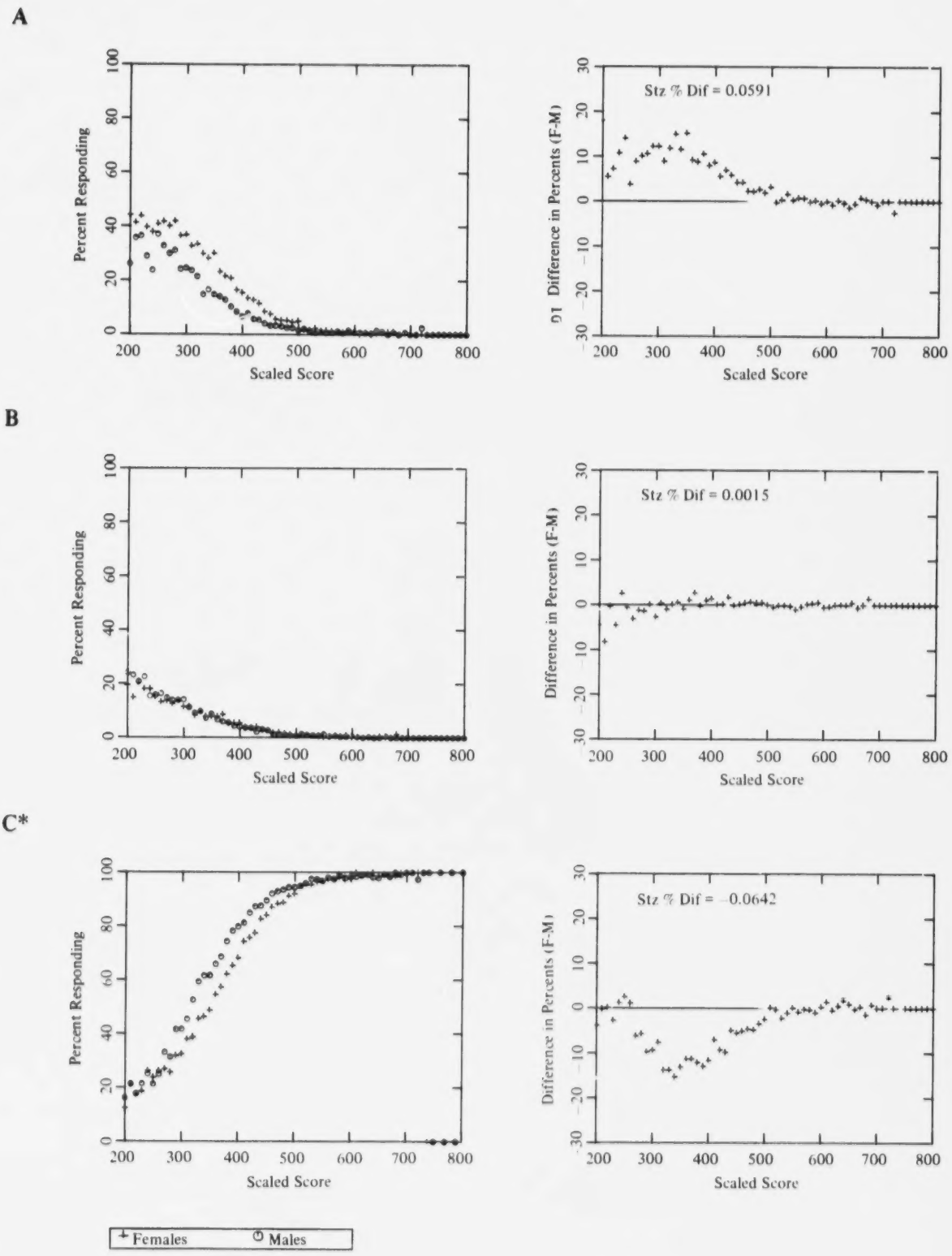
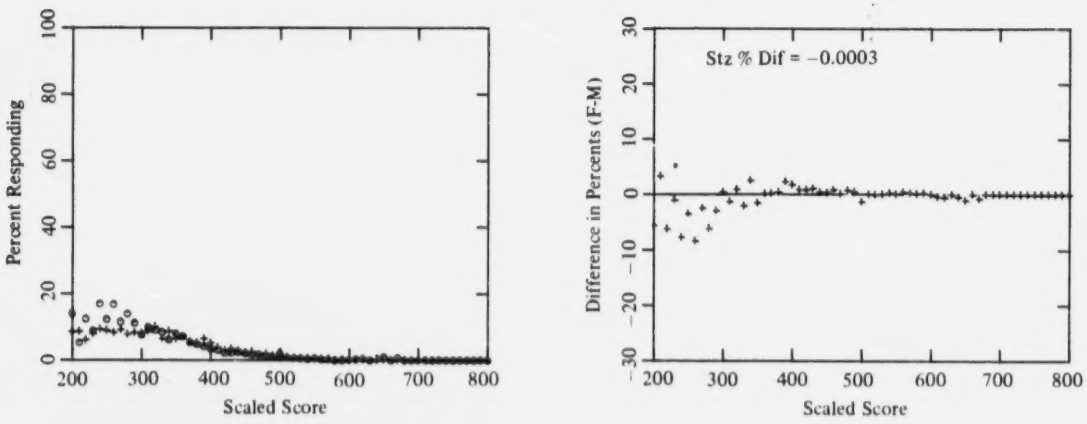
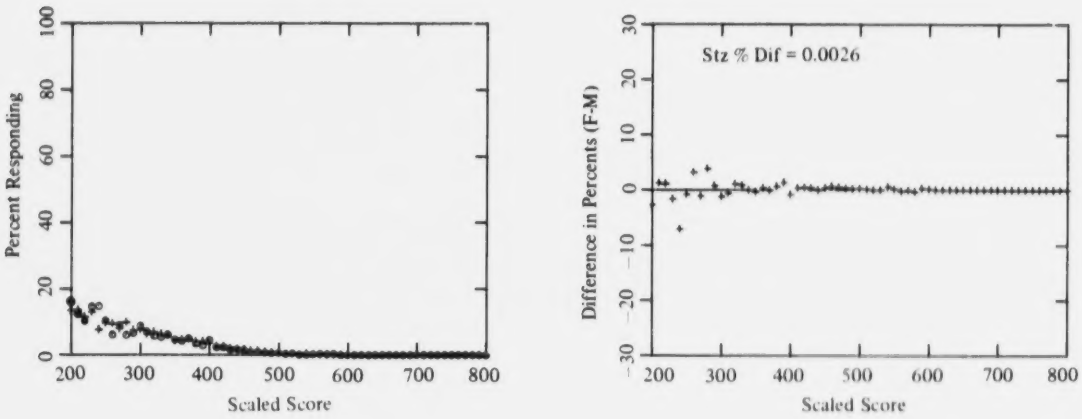


Figure B.4. Empirical Characteristic Functions for Sentence Completion Item 11 in Form 4, Verbal 2

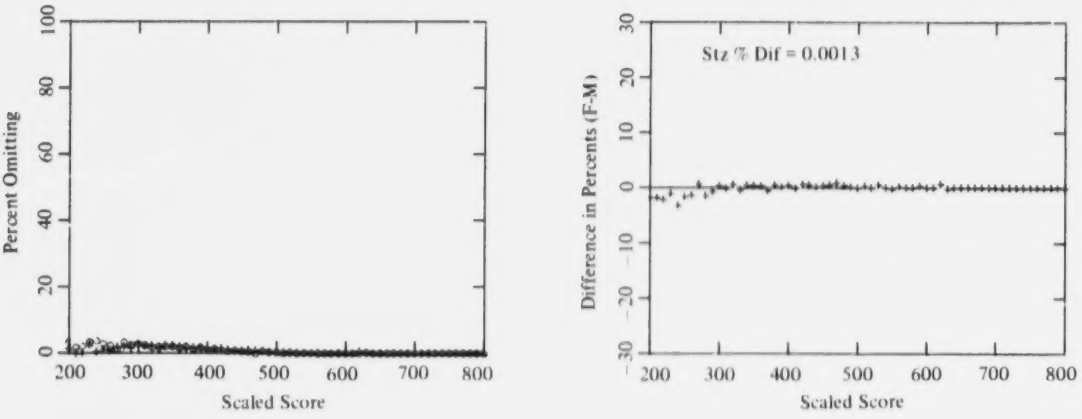
D



E



OMIT



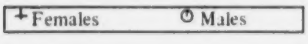


Figure B.4—*continued*

218277